





Digitized by the Internet Archive
in 2011 with funding from
MIT Libraries

<http://www.archive.org/details/bayesempiricalba277dumo>



**working paper
department
of economics**

Bayes and Empirical Bayes Methods
for Combining Cancer Experiments
in Man and Other Species

William H. DuMouchel*
and
Jeffrey E. Harris**

Number 277

February 1981

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**



Bayes and Empirical Bayes Methods
for Combining Cancer Experiments
in Man and Other Species

William H. DuMouchel*
and
Jeffrey E. Harris**

Number 277

February 1981

* Department of Mathematics, Massachusetts Institute of Technology.
Research supported by National Science Foundation Grant No. MCS-80-05483.

** Department of Economics, Massachusetts Institute of Technology.
Research supported by Public Health Service Research Grant No.
DA-02620 and Research Career Development Award No. DA-00072.

This paper is also issued as Technical Report No. 24, Department
of Mathematics, Massachusetts Institute of Technology.

BAYES AND EMPIRICAL BAYES METHODS FOR COMBINING
CANCER EXPERIMENTS IN MAN AND OTHER SPECIES

by

William H. DuMouchel*

and

Jeffrey E. Harris**

Massachusetts Institute of Technology

ABSTRACT

This paper offers a method for combining the results of diverse experiments when there is uncertainty about the relevance of some experiments to others. Within a Bayesian framework motivated by Lindley and Smith (1972), the method is used to assess human cancer risks from heterogeneous toxicological and epidemiological data. A distinction is drawn between the sampling error of each experiment and an error of relevance among experiments. The latter error reflects the uncertainty of interspecies extrapolations. It is shown how the experimental data, along with prior information on the credibility of such extrapolations, permits estimation of the human carcinogenic effects of various environmental emissions. A cross-validation method is proposed for selecting the most relevant subset among an array of experiments by eliminating those species or environmental agents which contribute most to the extrapolative error. Finally, other types of prior information on the relationships between experiments are incorporated into the analysis.

Key Words: CARCINOGENESIS; INTERSPECIES EXTRAPOLATION;
MUTAGENESIS; POLYAROMATIC HYDROCARBONS; LUNG
CANCER

* Research supported by National Science Foundation
Grant No. MCS-80-05483

** Research supported by Public Health Service Research
Grant No. DA-02620 and Research Career Development
Award No. DA-00072

1. INTRODUCTION

This paper offers a statistical method for combining the results of diverse experiments when there is uncertainty about the relevance of some experimental results to others. Our analysis is motivated by the increasing prominence of public policy problems in which decision makers call on multiple disciplines for advice. We seek an "enlargement of statistical techniques to encompass research programs rather than one at a time studies..." (Schneiderman, 1966).

We apply our method to the specific problem of assessing human cancer risk from an environmental agent when epidemiological data are imprecise or absent, but when precise toxicological studies in various species are available. This problem is more complicated than that posed by Cochran (1980), in which experiments of very similar design differed primarily in their date and location. In our problem, some experiments may be performed in vivo, while others are conducted in cell culture or in subcellular systems. Frequently, the experiments involve different compounds or mixtures of compounds. Interspecies comparisons are invariably required. Ideally, the exact relations among these experiments should be determined from fundamental advances in understanding the etiology of cancer. Our more modest goal here is to provide a statistical framework that permits scientists to combine the experimental results with their own prior judgments to reach quantitative conclusions. This objective is similar in spirit to those of Freireich et al. (1966), who compared the toxicity of anti-cancer agents in mouse, rat, hamster, dog, monkey, and man; Meselson and Russell (1977), who compared the mutagenic and carcinogenic potency of 14 compounds; and Crouch and Wilson (1979,1980), who examined the relative potencies of several chemical carcinogens in various pairs of species, most extensively in rats and mice.

The main idea behind our approach is to characterize precisely the different sources of variation among experiments. In our method, the results of each experiment are summarized by a single number, such as the slope of a dose-response relation. Each slope has an approximate standard error. These errors of measurement are assumed to be independent. The actual slopes, we hypothesize, lie near the response surface of an underlying regression model. Since some environmental agents may have distinctive effects in some species, this regression model necessarily entails some error. The critical factor linking the experiments is the scientist's a priori information on the exchangeability of these errors of interspecies extrapolation.

These ideas are formalized within a Bayesian framework similar to that of Lindley and Smith (1972). We assign prior distributions for the "hyperparameters" of the underlying regression model. Given these prior distributions and the experimental data, we compute the posterior distributions of the dose-response slopes. Empirical Bayes versions of our procedures are also presented.

In the next section, we pose a problem in the assessment of human lung cancer risk from a number of environmental emissions that contain polycyclic aromatic hydrocarbons. Section 3 formally develops our approach. Sections 4 and 5 then apply our method to the data. In Section 6, we offer a simple cross-validation procedure to assist in deciding which experiments are worth including. In Section 7, we discuss the case where a scientist has prior information on the relationships between experiments. The final section critically reviews our approach and suggests further lines of investigation.

Our calculations in this paper are illustrative. We do not propose here to draw conclusions about the public health significance of various ambient concentrations of pollutants. This would require a more thorough discussion

than space permits. Nor do we attach special significance to the dose-response models of carcinogenesis from which our data were derived. Harris (1981) has discussed the limitations of the use of such dose-response estimates in predicting excess cancer incidence from ambient population exposures.

2. THE PROBLEM

Table 1 displays the results of two types of carcinogenesis studies of two related environmental emissions, arranged in a 2x2 table. For each of the four experiments, three numbers are given: the observed slope of the dose-response relation; its coefficient of variation (i.e., the ratio of the standard error of the observed slope to its mean); and the natural logarithm of the observed slope. The first row of experiments represents the results of epidemiological studies of occupational exposures to coke oven emissions (Lloyd, 1971; Mazumdar et al., 1975; U.S. Environmental Protection Agency, 1979) and to roofing tar emissions (Hammond et al., 1976). The second row represents the results of skin tumor initiation experiments on the dichloromethane extracts of these emissions. The latter experiments were performed under identical conditions in the same laboratory, as part of the U.S. Environmental Protection Agency diesel emission research program (Nesnow et al., 1979; Huisingh et al., 1979). The slopes and their standard errors were estimated by maximum likelihood methods, as described in Harris (1981).

Our goal is to improve the precision of the estimated dose-response slopes for the human studies. The main question is how to use all of the data in Table 1 to achieve this objective.

One difficulty is immediately apparent. The dose-response slopes in man and mice are measured in different units. We might attempt to convert all the

TABLE 1.
2 x 2 Experimental Data Matrix

	Roofing Tar Emissions	Coke Oven Emissions	
Lung Cancer (Man)*	1.64	4.40	(slope)
	1.41	0.34	(coef.var.)
	0.49	1.48	(log slope)
Skin Tumor Initiation (Sencar Mice)**	0.54	2.10	
	0.04	0.04	
	-0.63	0.74	

*Increment in relative risk per 10^{-4} $\mu\text{g}/\text{m}$ extractable organics
x years.

**Papillomas/mouse per mg extract at 27 weeks.

experiments into common units, e.g., the incremental lifetime incidence of tumor per mg/kg body weight per day, or the age-specific probability of tumor per cumulative lifetime dose per unit body surface area. The choice of conversion factor, however, is hardly clear.

One way to circumvent this problem is to consider the relative potencies of the two environmental emissions in each species. Since the dose-response slopes in each row in Table 1 are measured in the same units, the ratios of the slopes are comparable unitless quantities. In fact, a natural hypothesis for combining these data is that the relative potency of the two emissions is preserved across the two biological systems.

The extent to which these data adhere to such an hypothesis can be ascertained in Figure 1, which depicts the means and standard errors of the dose-response slopes on a logarithmic scale. (The error bars correspond to the standard errors of the log slopes, which have been approximated by the coefficients of variation in Table 1.) On a log scale, the difference between coke oven slope and roofing tar slope in man is close to the corresponding difference in mice. To show this, we have also drawn the (weighted) least squares parallel lines on Figure 1, and the fit is good.

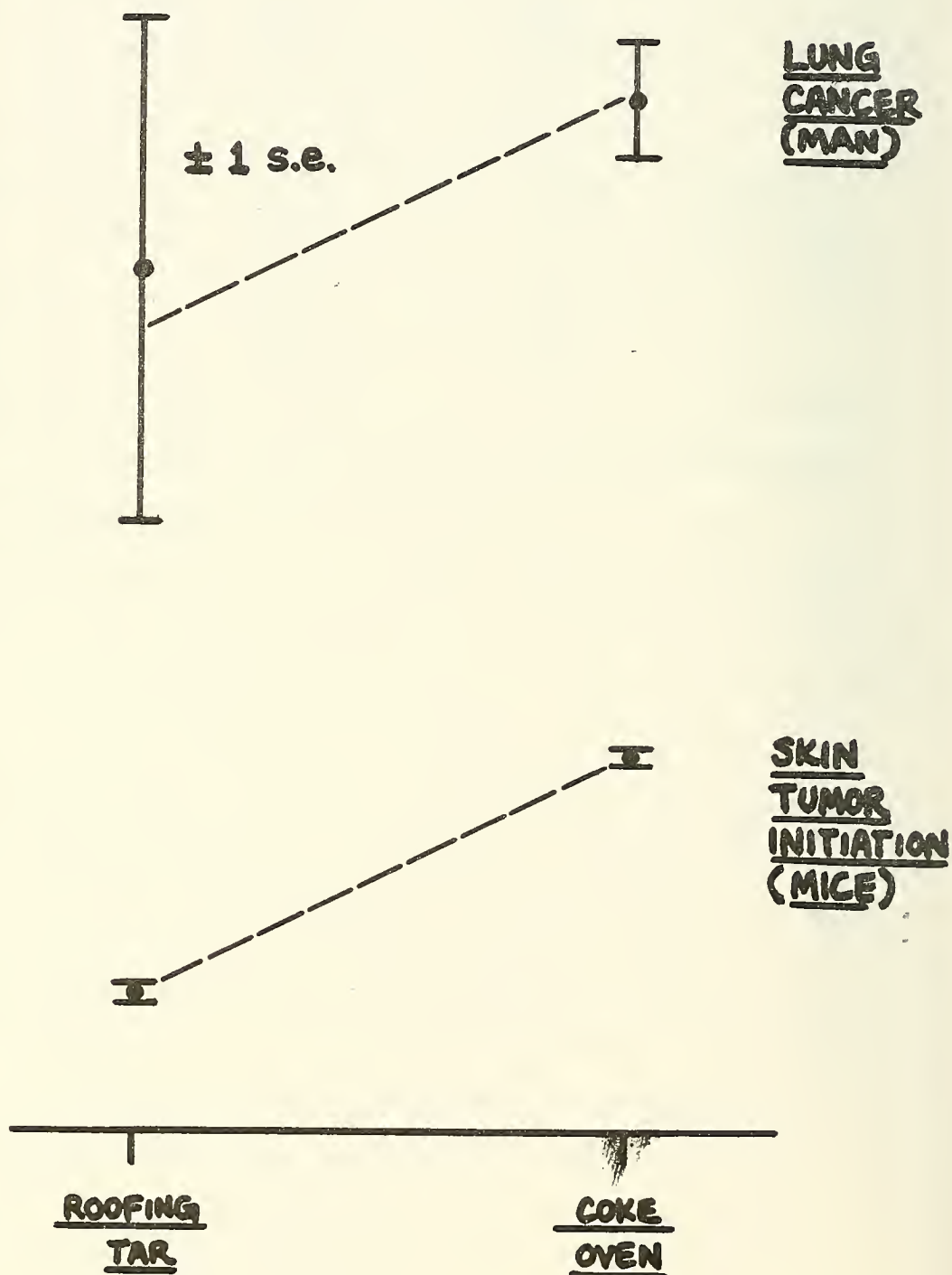
This result could be purely fortuitous. The standard errors of the epidemiological data, especially for roofing tar, are relatively large. But there is a deeper objection. The hypothesis that the relative carcinogenic potency of these two emissions is preserved across species ignores possible interspecies or interorgan differences in the distribution of particulates, the extractability of particulate-bound polyaromatic hydrocarbons, their clearance, metabolism, and genetic and other repair mechanisms. To claim that the totality of data in Table 1 provides more information about the human lung cancer risks from, say, roofing tar exposure than the roofing tar

increment in relative risk per 10^{-4} ng/m³
extractable organics x years

papillomas/mouse per mg
extract at 27 weeks

FIGURE 1

2x2 EXPERIMENTAL DATA MATRIX



epidemiological study alone is to maintain some degree of prior belief that these interspecies differences are not too large. The uncertainty inherent in such interspecies extrapolations clearly differs from the conventional sampling error of each experiment. If we are to use all of the data in Table 1 to estimate any one slope, then we must devise some measure of the extent of this extrapolative uncertainty.

Finally, there is the objection that the hypothesis of preserved relative potencies will not withstand other empirical comparisons. It is possible that such an hypothesis applies accurately only to the comparisons in Table 1, but not to other bioassays or to other environmental emissions. However, if we had no prior belief that the hypothesis should hold any more exactly for roofing tar and coke oven emissions than, say, for automotive particulate emissions or cigarette smoke, then any empirical comparisons that contradict the hypothesis would raise our uncertainty in the current extrapolation.

3. STATISTICAL MODEL

3.1 Notation and Assumptions.

Let y_{kl} be the logarithm of the estimated dose-response slope for the experiment in species k on environmental agent l . In the problem above, $k=1,2$ correspond to epidemiological studies in man and skin tumor initiation experiments in mice, respectively, while $l=1,2$ correspond to roofing tar emissions and coke oven emissions, respectively. The variables y_{kl} are presumed to be approximately normally distributed with mean θ_{kl} and known standard error c_{kl} . The assumptions of normality and known standard errors are not unreasonable, since each y_{kl} was a maximum likelihood estimate based on a relatively large experiment. The quantities θ_{kl} are the true log dose-response slopes, the primary parameters of interest.

We assume that each θ_{kl} has a symmetric prior distribution with mean value of the form

$$(3.1) \quad E[\theta_{kl} | \mu, \alpha_k, \gamma_l] = \mu + \alpha_k + \gamma_l,$$

where the hyperparameters $\{\mu, \alpha_k, \gamma_l\}$ represent the overall mean effect, species-specific effects, and emission-specific effects, respectively. In our Bayesian framework, these hyperparameters in turn have prior distributions. Equation (3.1) embodies the hypothesis that the relative potency of the two emissions is on average preserved across species. Moreover, the relative potencies are a priori just as likely to be larger for one species than for the other. The various θ_{kl} are measured in different units, since they are the logs of dose-response slopes for quite different dose-response experiments. The additive model (3.1) is meaningful, however, so long as $(\theta_{11} - \theta_{12}) - (\theta_{21} - \theta_{22})$ is a dimensionless quantity, a condition satisfied in our problem above. In that case, the units of measurement for μ, α_k , and γ_l can be chosen so that the quantities

$$\delta_{kl} = \theta_{kl} - \mu - \alpha_k - \gamma_l$$

are similarly dimensionless. Each δ_{kl} is a species-emission interaction effect, measuring the amount (on a log scale) by which the experiment in species k on emission l deviates from the constant relative potency hypothesis.

Conditional on the value of another hyperparameter σ , we further assume that the δ_{kl} are independently distributed a priori as $N(0, \sigma^2)$. Under this critical assumption, the interaction effects δ_{kl} are a priori exchangeable

(de Finetti, 1964). That is, we have no prior information that a deviation of a given magnitude from the constant relative potency model is more likely in one experiment than in any other.

We take care here to elucidate the precise meaning of this form of prior information. We recognize that the mechanisms of carcinogenesis may vary considerably among agents or species. Quite different metabolic pathways may be involved. The number of stages in expression of tumor may differ. Any variation that is distinctive to a particular agent in a particular species could result in a marked deviation from our additive model. The exchangeability hypothesis does not exclude the possibility of such deviations. It merely states that we cannot identify a priori which entry in our two-way table of experiments is likely to have the largest deviation.

The hyperparameter σ measures our belief in the degree of accuracy of the equal relative potency model. A value of $\sigma=0.05$, for example, implies that within one normal standard deviation, i.e., with probability 0.68, the additive model is accurate to within an absolute error of 0.05; or equivalently, each dose-response slope conforms to the underlying equal relative potency model to a multiplicative factor of $\exp(0.05) \approx 1.05$. A prior belief that σ is of this magnitude thus implies a relatively high degree of confidence in the underlying model. On the other hand, a value of $\sigma = 5$ implies that with probability 0.68, each dose-response slope conforms to the underlying model to a multiplicative factor of $\exp(5) \approx 150$. A belief that σ is of this magnitude implies much less faith that the experiments can be profitably combined.

We now generalize beyond the 2×2 case considered above, letting $\{ y_{kl} \pm c_{kl} ; k=1, \dots, K ; l=1, \dots, L \}$ be a set of experimental observations on the log dose-response slopes for K species and L environmental agents. We also

admit the possibility that some y_{kl} are missing from a set of contemplated experiments. Except where otherwise noted below, we assume that the set of available experiments is connected, in the sense that any available experiment (k, l) can be reached from any other available experiment (k', l') by a series of moves from one available experiment to another in which each move is along a single row or column. Conditional on σ , the observed y_{kl} are generated by the linear model

$$(3.2) \quad y_{kl} = \mu + \alpha_k + \gamma_l + \delta_{kl} + \varepsilon_{kl},$$

where the three sets of variables $\{\mu, \alpha_k, \gamma_l\}$, $\{\delta_{kl}\}$, and $\{\varepsilon_{kl}\}$ are independent a priori, with the δ_{kl} i.i.d. $N(0, \sigma^2)$ and the ε_{kl} independent $N(0, c_{kl}^2)$. Following the usual general linear model formulation, we further replace the expressions $\mu + \alpha_k + \gamma_l$ in (3.2) by $X\beta$, where β is a column vector of hyperparameters and X is an appropriately chosen design matrix. Of the $K+L+1$ hyperparameters in $\{\mu, \alpha_k, \gamma_l\}$, at most $K+L-1$ are independently estimable in the classical sense. So long as we use an informative full rank prior distribution on all $K+L+1$ hyperparameters, no restrictions on these hyperparameters are necessary in our Bayesian framework. In other cases, however, particularly when a diffuse prior distribution is employed, we shall assume that β corresponds only to the independently estimable components of $\{\mu, \alpha_k, \gamma_l\}$ and that X is the corresponding full rank design matrix.

Finally, we assume that β is a a priori multivariate normal with mean vector b and covariance matrix V , and that σ has a prior distribution π with density $\pi(\sigma)$. Now let $i = 1, \dots, n$ index the experiments, replacing the paired indices (k, l) . Let m be the rank of X , that is, the

number of independently estimable elements of $\{\mu, \alpha_k, \sigma_k\}$. Let y, θ, δ , and ε be $n \times 1$ column vectors replacing $\{y_{k\ell}\}$, $\{\theta_{k\ell}\}$, $\{\delta_{k\ell}\}$, and $\{\varepsilon_{k\ell}\}$, respectively. Let I be the $n \times n$ identity matrix, and let C be $\text{diag}(c_1^2, \dots, c_n^2)$. Our model can be formulated generally as

$$Y = X\beta + \delta + \varepsilon, \text{ where } \theta = X\beta + \delta, \text{ and}$$

$$(3.3a) \quad \sigma \sim \pi,$$

$$(3.3b) \quad \beta \sim N(b, V),$$

$$(3.3c) \quad (\theta | \beta, \sigma) \sim N(X\beta, \sigma^2 I),$$

$$(3.3d) \quad (Y | \theta) \sim N(\theta, C).$$

This model possesses a hierarchical structure similar to that formulated by Lindley and Smith (1972). The experimental data Y and C , as well as b , V , and the distribution π , are assumed to be known. The choice of a prior distribution π is left unspecified until Section 4. We note here that there is no advantage or compelling reason to choose the inverse chi-squared prior distribution for σ , as proposed by Smith (1973a). The choice of b and V is more complicated, and will be considered in detail below. (Readers who are less interested in the mathematical details of estimation may wish to skim the remainder of Section 3 and resume in earnest at Section 4.)

3.2 Bayes Estimates. Informative Prior on β .

Let us suppose that a scientist has prior information on β , which he expresses through his choice of b and V . Such choices could be made directly, as we shall illustrate in Section 7, or by indirect elicitation, as in the method of Kadane et al. (1980).

Estimation of the parameters now proceeds by straightforward Bayesian methodology. First, we compute the marginal distribution of the data given the hyperparameter σ . From (3.3),

$$(3.4) \quad (Y|\sigma) \sim N(Xb, C + \sigma^2 I + XVX').$$

In our Bayesian framework, (3.4) can be regarded as the likelihood of σ for a priori given values of b and V . The posterior distribution of σ is therefore

$$(3.5) \quad \pi(\sigma|Y) \propto \pi(\sigma) |C + \sigma^2 I + XVX'|^{-1/2} \exp\{-\frac{1}{2}(Y-Xb)' [C + \sigma^2 I + XVX']^{-1} (Y-Xb)\}$$

where $|A|$ is the determinant of A . For future reference, we define the posterior expectation of σ^2 as

$$\sigma^{*2} = \int_0^{\infty} \sigma^2 \pi(\sigma|Y) d\sigma,$$

which can be interpreted as the approximate risk in estimating a particular θ by $X\beta$ under squared error loss.

Now consider the posterior distribution of β . Denoting the posterior density by $f(\beta|Y)$, we have from (3.3)

$$(3.6) \quad f(\beta|Y) = \int_0^{\infty} f(\beta|Y, \sigma) \pi(\sigma|Y) d\sigma,$$

where $f(\beta|Y, \sigma)$ is the multivariate normal density $N(\tilde{\beta}, \tilde{V})$, and

$$(3.7) \quad \tilde{\beta} = \tilde{V}[X'(C + \sigma^2 I)^{-1} Y + V^{-1} b],$$

$$\tilde{V} = [X'(C + \sigma^2 I)^{-1} X + V^{-1}]^{-1}.$$

The posterior distribution of β is a mixture of multivariate normal distributions with mixing probabilities given by (3.5), the posterior density of σ . Equations (3.7) are derived by the familiar rule for computing posterior distributions for the normal data, normal prior, known variance case. (See, e.g., Raiffa and Schlaifer (1968).) With $W = (C + \sigma^2 I)^{-1}$ known, the least squares estimator $\hat{\beta} = (X'WX)^{-1} X'WY$ has mean β and precision matrix $(X'WX)$ (where, for the sake of this intuitive argument, X here is necessarily of full rank). Moreover, the prior distribution of β has mean b and precision V . The rule for computing the posterior distribution of β is to weight the least squares estimate and the prior mean by their precisions, with the precision of the result equal to the sum of these precisions. In the analysis below, we shall be interested in the fitted values $X\beta$ from the underlying constant relative potency hypothesis. The posterior density of $X\beta$ is the corresponding mixture of multivariate normal densities $N(X\tilde{\beta}, X\tilde{V}X')$, where the mixing probabilities are still $\pi(\sigma|Y)$ and $\tilde{\beta}$ and $\tilde{V}$ are defined in (3.7).

Consider, finally, the estimation of θ . If $g(\theta|Y)$ is the posterior density of θ , we have

$$(3.8) \quad g(\theta|Y) = \int_0^\infty g(\theta|Y, \sigma) \pi(\sigma|Y) d\sigma,$$

where $g(\theta|Y, \sigma)$ is the multivariate normal density $N(\tilde{\theta}, \tilde{C})$, and

$$(3.9) \quad \begin{aligned} \tilde{\theta} &= \tilde{C}[C^{-1}Y + (XVX' + \sigma^2 I)^{-1}Xb], \\ \tilde{C} &= [C^{-1} + (XVX' + \sigma^2 I)^{-1}]^{-1}. \end{aligned}$$

The posterior distribution of θ is similarly a mixture of normal distributions. The means and covariances of these normal distributions, given by (3.9), are derived in a manner analogous to (3.7), where, by (3.3), $Y|\theta, \sigma \sim N(\theta, C)$ and $\theta|\sigma \sim N(Xb, XVX' + \sigma^2 I)$. Each $\tilde{\theta}$ is a weighted average of the original data Y and the corresponding prior prediction Xb from the underlying constant relative potency model, where the weights are the corresponding precisions.

3.3 Bayes Estimates. Diffuse Prior on β .

Calculation of the posterior distributions (3.5), (3.6), and (3.8) requires us to specify the mean b and the covariance matrix V of the prior distribution of β . In many situations, however, information about the hyperparameters β will be extremely vague. That is, the prior covariance matrix V will be large. To investigate such cases in detail, we first need the following lemma.

Lemma: Let U be an $n \times m$ matrix of rank $m < n$, I be the $n \times n$ identity matrix, and t be a scalar. Then as $t \rightarrow \infty$,

$$(3.10) \quad (I + tUU')^{-1} = I - U(U'U)^{-1}U' + t^{-1}(UU')^+ + O(t^{-2}),$$

$$(3.11) \quad |I + tUU'| = t^m |U'U| [1 + t^{-1} \text{tr}(UU')^+ + O(t^{-2})],$$

where A^+ is the Moore-Penrose pseudo-inverse of A .

Proof: The $n \times n$ matrix UU' , which has rank m , can be represented as

$$UU' = \sum_{j=1}^m \lambda_j u_j u_j',$$

where $\{\lambda_j\}$ are the strictly positive characteristic roots of UU' and $\{u_j\}$ are the corresponding characteristic vectors. The $n \times n$ identity matrix can be represented as

$$I = \sum_{j=1}^m u_j u_j' + \sum_{j=m+1}^n v_j v_j' ,$$

where the unit vectors $\{v_j\}$ are all orthogonal to the characteristic vectors $\{u_j\}$. Combining these two expressions, we have

$$I + tUU' = \sum_{j=1}^m (1+t\lambda_j) u_j u_j' + \sum_{j=m+1}^n v_j v_j' .$$

Now

$$(I+tUU')^{-1} = \sum_{j=1}^m (1+t\lambda_j)^{-1} u_j u_j' + \sum_{j=m+1}^n v_j v_j' .$$

As $t \rightarrow \infty$,

$$\sum_{j=1}^m (1+t\lambda_j)^{-1} u_j u_j' = t^{-1} \sum_{j=1}^m \lambda_j^{-1} u_j u_j' + O(t^{-2}).$$

Equation (3.10) now follows from our recognition that $\sum_{j=m+1}^n v_j v_j'$ is the orthogonal projection operator which maps R^n onto the subspace of R^n orthogonal to the columns of U (namely $I - U(U'U)^{-1}U'$), while $\sum_{j=1}^m \lambda_j^{-1} u_j u_j'$ defines the pseudo-inverse. Similarly

$$|I+tUU'| = \prod_{j=1}^m (1+t\lambda_j) = t^m \prod_{j=1}^m \lambda_j (1+t^{-1} \sum_{j=1}^m \lambda_j^{-1} + O(t^{-2})).$$

Equation (3.11) follows from our recognition that $|U'U| = \prod_{j=1}^m \lambda_j$ and $\text{tr}(UU')^+ = \sum_{j=1}^m \lambda_j^{-1} \times$

We now have the following result.

Proposition: If the nonsingular covariance matrix V is replaced by tV , where t is a scalar, then as $t \rightarrow \infty$:

(a) The posterior density of σ approaches

$$(3.12) \quad \pi(\sigma|Y) \propto \pi(\sigma) |W|^{1/2} |X'WX|^{-1/2} \exp\{-\frac{1}{2}Y'SY\},$$

where $W = (C + \sigma^2 I)^{-1}$ and $S = W - WX(X'WX)^{-1}X'W$.

(b) The posterior density of β approaches $f(\beta|Y) = \int_0^\infty f(\beta|Y, \sigma) \pi(\sigma|Y) d\sigma$, where $f(\beta|Y, \sigma)$ is multivariate normal $N(\tilde{\beta}, \tilde{V})$ and

$$(3.13) \quad \begin{aligned} \tilde{\beta} &= \tilde{V}X'(C + \sigma^2 I)^{-1}Y \\ \tilde{V} &= [X'(C + \sigma^2 I)^{-1}X]^{-1}. \end{aligned}$$

(c) The posterior density of θ approaches $g(\theta|Y) = \int_0^\infty g(\theta|Y, \sigma) \pi(\sigma|Y) d\sigma$, where $g(\theta|Y, \sigma)$ is multivariate normal $N(\tilde{\theta}, \tilde{C})$ and

$$(3.14) \quad \begin{aligned} \tilde{\theta} &= \tilde{C}C^{-1}Y = (I + CR)^{-1}Y \\ \tilde{C} &= (C^{-1} + R)^{-1}, \end{aligned}$$

where $R = \sigma^{-2}[I - X(X'X)^{-1}X']$.

Proof: (a) Note that the quadratic form $Y'SY$ in (3.12) is the sum of

squared residuals of the weighted least squares regression of Y on the columns of X , where the weights are the diagonal elements of W . That the quadratic form $(Y-Xb)'[C+\sigma^2 I+XVX']^{-1}(Y-Xb)$ in (3.5) reduces to $Y'SY$ in (3.12) is a result of expansion formula (3.10), where we set $U = W^{1/2}XV^{1/2}$ and $S = W^{1/2}(I-U(U'U)^{-1}U')W^{1/2}$. That the determinant $|C+\sigma^2 I+XVX'|^{-1/2}$ in (3.5) becomes proportional to $|W|^{1/2}|X'WX|^{-1/2}$ in (3.12) is a result of expansion formula (3.11) under the same definition of U . (As noted in Section 3.1, we assume here a parametrization in which X is of full rank.)

(b) Expressions (3.13) follow from our setting $V^{-1} = 0$ in expressions (3.7).

(c) That expression (3.9) reduces to (3.14) is a result of the expansion formula (3.10), where we set $U = XV^{1/2}$ in order to evaluate the terms $(XVX'+\sigma^2 I)^{-1}$ in (3.9). We note also that equation (3.14) is a special case of equation (A1) of Smith (1973a). *

3.4. Empirical Bayes Estimates.

In the analysis below, we shall also consider empirical Bayes approaches to estimating θ . In these alternative methods, we retain the prior distribution $\theta|\beta, \sigma \sim N(X\beta, \sigma^2 I)$ as specified in (3.3c), but use the data Y itself to construct the prior distributions on σ and β .

Several options are available. (As Dempster (1980) notes, "there is no such thing as the empirical Bayes estimator.") First, we could estimate both σ and β from the data Y , by maximum likelihood or other methods, and then assume that the entire prior density for σ is concentrated at the estimate $\hat{\sigma}$ and the entire prior density of β is concentrated at the estimate $\hat{\beta}$. For a given σ , the maximum likelihood estimate of β is the least squares estimate

$$\hat{\beta}_{MLE}(\sigma) = (X'(C + \sigma^2 I)^{-1} X)^{-1} X'(C + \sigma^2 I)^{-1} Y.$$

As a function of σ , the concentrated likelihood, evaluated at $\beta = \hat{\beta}_{MLE}$, is proportional to

$$\hat{L}(\sigma) = |W|^{1/2} \exp\{-\frac{1}{2} Y'(W - WX(X'WX)^{-1} X'W)Y\}$$

where, again, $W = (C + \sigma^2 I)^{-1}$. If we denote $\hat{\sigma}_{MLE}$ as the value of σ maximizing this likelihood function, then the resulting empirical Bayes posterior distribution of θ is $N(\hat{\theta}_{MLE}, \hat{C}_{MLE})$, where

$$(3.15) \quad \begin{aligned} \hat{\theta}_{MLE} &= \hat{C}_{MLE} [C^{-1} Y + \hat{\sigma}_{MLE}^{-2} X \hat{\beta}_{MLE}(\hat{\sigma}_{MLE})] \\ \hat{C}_{MLE} &= [C^{-1} + \hat{\sigma}_{MLE}^{-2} I]^{-1}. \end{aligned}$$

This estimate treats β as if it were fixed and known a priori, even though the estimate $\hat{\beta}_{MLE}$ is used.

Alternatively, we could assume a diffuse prior on β and estimate only σ from the data Y . In this case, the appropriate likelihood function for σ is equation (3.12) with the prior density $\pi(\sigma)$ omitted, that is,

$$L^*(\sigma) = |W|^{1/2} |X'WX|^{-1/2} \exp\{-\frac{1}{2} Y'SY\}.$$

Let $\hat{\sigma}_{EB}$ be the value of σ that maximizes $L^*(\sigma)$. The corresponding empirical Bayes posterior distribution treats $\sigma = \hat{\sigma}_{EB}$ as if it were known with certainty, and so, by equation (3.14), is $N(\hat{\theta}_{EB}, \hat{C}_{EB})$, where

$$(3.16) \quad \begin{aligned} \hat{\theta}_{EB} &= [I + \hat{\sigma}_{EB}^{-2} C (I - X(X'X)^{-1}X')]^{-1} Y \\ \hat{C}_{EB} &= [C^{-1} + \hat{\sigma}_{EB}^{-2} (I - X(X'X)^{-1}X')]^{-1}. \end{aligned}$$

It is interesting to note that in the case where σ is known, Smith (1973b) shows that if we replace $\hat{\sigma}_{MLE}$ and $\hat{\sigma}_{EB}$ by σ , then $\hat{\theta}_{MLE} = \hat{\theta}_{EB}$. It is clear from comparison of (3.15) with (3.16) that if $\hat{\sigma}_{MLE} = \hat{\sigma}_{EB}$, then $\hat{C}_{MLE} \leq \hat{C}_{EB}$ (in the sense that $\hat{C}_{MLE} - \hat{C}_{EB}$ is non-negative definite). Moreover, if $\hat{\sigma}_{MLE} < \hat{\sigma}_{EB}$, then $\hat{C}_{MLE} < \hat{C}_{EB}$. But in fact $\hat{\sigma}_{MLE}$ is less than $\hat{\sigma}_{EB}$, since they maximize $\hat{L}(\sigma)$ and $L^*(\sigma)$, respectively, and since the ratio $\hat{L}(\sigma)/L^*(\sigma) = |X'(C + \sigma^2 I)^{-1}X|^{1/2}$ can be shown to be a decreasing function of σ . Since L^* is the product of $\hat{L}$ and an increasing function of σ , its maximum will occur later than that of $\hat{L}$. The assumption that $\beta = \hat{\beta}_{MLE}$ with certainty leads to a smaller posterior variance for θ than the diffuse prior for β . Therefore, to the extent that β is a priori uncertain, the variance $\hat{C}_{MLE}$ is inappropriately small. In the results below, we shall therefore report the Empirical Bayes estimate (3.16).

Finally, if we wish to avoid the computational burden in determining $\hat{\sigma}_{MLE}$ or $\hat{\sigma}_{EB}$, we could begin with $\hat{\beta}_{MLE}(0) = (X'C^{-1}X)^{-1}X'C^{-1}Y$. The residual sum of squares for this estimate, $RSS = \sum_{i=1}^n (y_i - x_i'\hat{\beta})^2 / c_i^2$, has expectation $E[RSS] = n - m + \sigma^2 [\sum_{i=1}^n c_i^{-2} - \text{tr } C^{-1}X(X'C^{-1}X)^{-1}X'C^{-1}]$, which suggests the estimate

$$(3.17) \quad \hat{\sigma}_{RSS}^2 = [RSS - (n-m)] / [\sum_{i=1}^n c_i^{-2} - \text{tr } C^{-1}X(X'C^{-1}X)^{-1}X'C^{-1}]$$

where we take $\hat{\sigma}_{RSS}^2 = 0$ if $RSS < n-m$. (The exact value of $E[RSS]$ was derived for us by H. Chernoff, whose proof is omitted.) In the results below, we shall also report the empirical Bayes estimate of θ when $\hat{\sigma}_{RSS}$ is

substituted for $\hat{\sigma}_{EB}$ in (3.16).

4. THE 2x2 CASE

In the next three sections of this paper, we shall assume a diffuse prior on the vector of hyperparameters β . To be sure, a scientist may have prior information on the composition of each emission, the carcinogenic activities of its constituents, their possible synergistic interactions, their bioavailability, etc. Similarly, a scientist may have prior information on the sensitivity of the mouse skin tumor initiation model in comparison to the human respiratory tract. Our impression, however, is that this type of information is not yet sufficiently refined to offer much help in specifying a precise prior on β . We recognize that the use of improper priors may involve certain marginalization paradoxes. Discussion of these potential difficulties is deferred to Section 7. At this point, we note that the analysis of the next three sections was repeated under the assumption of a proper prior for β with $V = 10^4 I$. The results of all reported quantities were unchanged up to the number of decimals presented.

Devising a prior distribution for the critical hyperparameter σ is another matter. Perfect extrapolation from mouse to man or from one environmental agent to another is clearly quite unlikely. To claim that the various experiments in Table 1 are totally irrelevant to each other is likewise too strong. The answer lies somewhere in between.

It seems reasonable to suspect that within a range of one normal standard deviation, i.e., with probability 0.68, the underlying constant relative potency model could be accurate within a multiplicative factor of $\exp(1) \approx 2.7$, or even $\exp(0.5) \approx 1.6$. To suspect that, with probability 0.68, the model could be accurate within a factor of $\exp(0.05) \approx 1.05$ is more

controversial. One of us, in fact, found the possibility of significant interspecies differences in particulate distribution, extraction, clearance, metabolism, etc. so compelling that an a priori error factor of 1.05 seemed unimaginable. On the other hand, we both felt that it would be inappropriate to attach a uniform distribution to σ , since there is uncertainty even in the order of magnitude of error. To articulate our differences, and agreements, we formulated two prior distributions on σ .

Prior A: $\log \sigma$ uniformly distributed on the interval

$$0.05 \leq \sigma \leq 5; \text{ and}$$

Prior B: $\log \sigma$ uniformly distributed on the interval

$$0.5 \leq \sigma \leq 5.$$

These distributions somewhat artificially attach zero probability mass outside the specified intervals $[0.05, 5]$ and $[0.5, 5]$. As we shall see shortly, however, this restriction does not significantly affect our main conclusions. Giri (1970) has employed a uniform distribution on $\log \sigma$ for $0 < \sigma < \infty$ in his Bayesian model for two-way ANOVA. However, we prefer the use of a proper prior distribution because it compels us to face the task of assessing our beliefs. Priors A and B retain the feature that, within the relevant intervals, the posterior density of $\log \sigma$ will be proportional to the likelihood function.

In order to simplify the computations, we shall evaluate these prior distributions, and therefore the posterior distribution $\pi(\sigma|Y)$, only at discrete points in the relevant intervals, equally spaced on the $\log$ scale. This means that the posterior distributions of $X\beta$ and θ will be finite mixtures of normal distributions.

Figure 2 displays the posterior densities $\pi(\sigma|Y)$ calculated from (3.12) for our two distinct priors. For the 2×2 data matrix in Table 1, the form of the posterior distribution of σ is clearly sensitive to the prior distribution that is assumed. In the interval $0.05 \leq \sigma \leq 0.5$, in particular, the likelihood function is relatively flat. Yet the maximum likelihood estimate of σ is zero.

This finding is reflected in Table 2, which shows selected statistics of the posterior distributions of $X\beta$ and θ for the epidemiological studies, based upon the two prior distributions of σ . Also shown are the results for the empirical Bayes estimate corresponding to (3.16). (Both $\hat{\sigma}_{EB}$ and $\hat{\sigma}_{RSS}$ were zero in this case.)

Although the posterior distribution of θ is a mixture of multivariate normals (recall (3.8)), the resulting marginal distributions did not in fact deviate substantially from normality. If $\theta_i^* = E[\theta_i|Y]$ and if c_i^* is the standard deviation of $\theta_i|Y$, then the tail probabilities

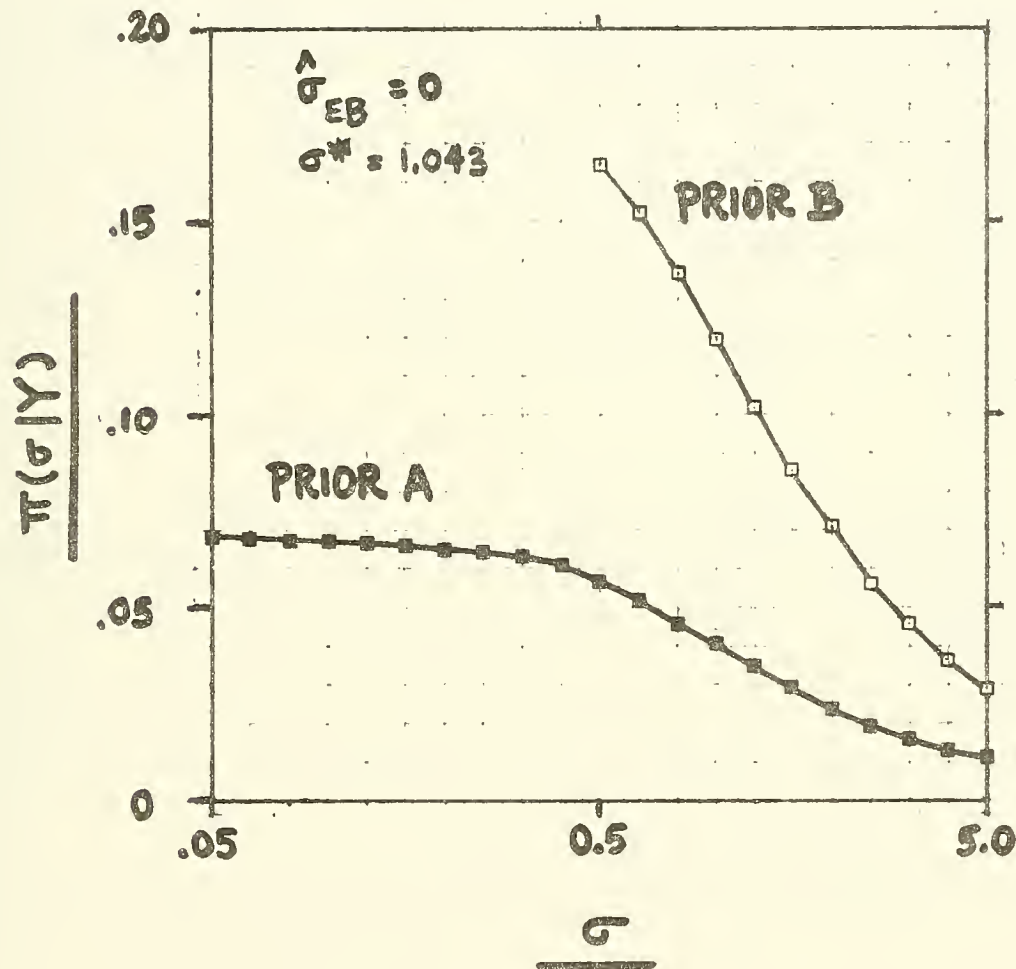
$$\Pr\{\theta_i \geq \theta_i^* + 2.326c_i^*|Y\},$$

$$\Pr\{\theta_i \leq \theta_i^* - 2.326c_i^*|Y\},$$

do not deviate substantially from the value of 0.01 predicted for the normal density. Hence, the mean and standard deviation adequately characterize the marginals of the posterior distribution of θ .

Because the original coke oven data were relatively precise, the means and standard deviations of the posterior distributions of the coke oven log slope do not differ much from the original values of y and c . For the roofing tar log slope, however, the precision of the posterior distribution depends critically on the estimate used. Since prior A admits the possibility

FIGURE 2: Posterior Densities of $\log \sigma$
2 x 2 EXPERIMENTAL DATA MATRIX



Prior A: $\log \sigma$ uniform on $0.05 \leq \sigma \leq 5.0$

Prior B: $\log \sigma$ uniform on $0.5 \leq \sigma \leq 5.0$

TABLE 2.

Bayes and Empirical Bayes Estimates of Log Slopes
For Lung Cancer Risk in Man
2 x 2 Data Matrix

Environmental Emission	Mean	Stand. Dev.	Poster. Mean $X\beta$	Lower Tail	Upper Tail
Roofing Tar					
Original Data	0.495	1.415			
θY (Prior B)	0.365	1.152	0.304	0.011	0.013
θY (Prior A)	0.229	0.788	0.205	0.015	0.025
θY (Empirical Bayes)	0.135	0.337	0.135		
Coke Oven					
Original Data	1.482	0.341			
θY (Prior B)	1.489	0.338	1.550	0.010	0.010
θY (Prior A)	1.497	0.334	1.522	0.010	0.010
θY (Empirical Bayes)	1.502	0.331	1.502		

$\theta|Y$ (Empirical Bayes) assumes $\pi(\sigma)$ concentrated at $\hat{\sigma}_{EB} = 0$, and diffuse prior on β .

Lower Tail = $\Pr\{\theta_i \leq \theta_i^* - 2.326c_i^* | Y\}$,
Upper Tail = $\Pr\{\theta_i \geq \theta_i^* + 2.326c_i^* | Y\}$,
where θ_i^* and c_i^* are the posterior mean and standard deviation of θ_i .

of lower values of σ , the corresponding posterior distribution has a smaller standard deviation. For the empirical Bayes estimate, which in this case assumes that the underlying constant relative potency model holds exactly, the only sources of variance in the posterior distribution of θ are the original sampling errors.

The posterior mean values of θ are very close to the corresponding posterior mean values of $X\beta$. That is, the posterior expectations of the model residuals δ are small. The posterior variances of these residuals, however, are not so small. Although the variance of each posterior residual $\delta_i | Y$ depends in part on the precision of the original data, that component of the variance due purely to the underlying model is

$$(4.1) \quad \sigma^{*2} = E[\sigma^2 | Y],$$

which for prior A in this case is 1.088. In effect, if we were to use these data to predict δ for another experiment yet to be performed, the standard deviation of δ , under Prior A, would be 1.04. Under Prior B, the standard deviation of δ would be 1.76, despite the fact that the empirical Bayes estimate of σ is $\hat{\sigma}_{EB} = 0$. (It is straightforward to show that $\delta | Y$ is likewise a mixture of normals, each of which has covariance matrix of the form $\sigma^2 I + D$, where D vanishes as C^{-1} approaches 0.)

Little credence, we conclude, can be attached to the apparently close fit of the data in Table 1 to the underlying constant relative potency model. Any scientist who objects that the data are just "too good to be true" makes a legitimate claim based on his prior belief that such extrapolative models are unlikely to be so accurate. The extent to which the totality of data in Table 1 refines the precision of the estimated human lung cancer risk is, in effect,

a matter of prior opinion.

The problem with the 2×2 case, it appears, is that we don't have enough precise experiments. The sampling errors for the skin tumor initiation data in mice are so small that the model is fitted, in effect, to the mouse data. The predicted relative potencies for the human lung cancer risks merely adjust to the more precise non-human results. If we are to learn any more about the extent to which these experiments can be combined, then we need additional precise experiments. We now proceed in this direction.

5. THE 3×3 CASE

Table 3 is an augmented version of Table 2. In addition to human lung cancer epidemiological studies and skin tumor initiation experiments in mice, we have included experiments on the enhancement of viral oncogenic transformation in Syrian hamster embryo (SHE) cells (Casto et al., 1979). In addition to studies on roofing tar and coke oven emissions, we have included experiments on the dichloromethane extracts of particulate emissions from one light duty diesel engine.

Except for the epidemiological studies, experiments appearing in the same row were, as above, performed under identical conditions in the same laboratory. The new slopes and standard errors were, as above, estimated by maximum likelihood methods, as described in Harris (1981). No epidemiological study of the human lung cancer risks from exposure to light duty diesel engine exhaust was available. Although the corresponding cell is left empty, we note that the set of available experiments is connected, as defined in Section 3.1 above.

The results in Table 3 clearly reveal inconsistencies in the constant

TABLE 3.

3 x 3 Experimental Data Matrix

	Roofing Tar Emissions	Coke Oven Emissions	Diesel Engine Emissions
Lung Cancer (Man)	1.64 1.41 0.49	4.40 0.34 1.48	slope coef.var. log slope
Skin Tumor Initiation (Sencar Mice)	0.54 0.04 -0.63	2.10 0.04 0.74	0.53 0.04 -0.64
Enhancement of Viral Transformation (SHE Cells)*	2.07 0.18 0.73	0.86 0.10 -0.15	0.65 0.15 -0.44

*Transformations/ 2×10^6 cells per $\mu\text{g/ml}$ extract .
Units for other rows as in Table 1.
There are no data for lung cancer risk of diesel engine emissions in man.

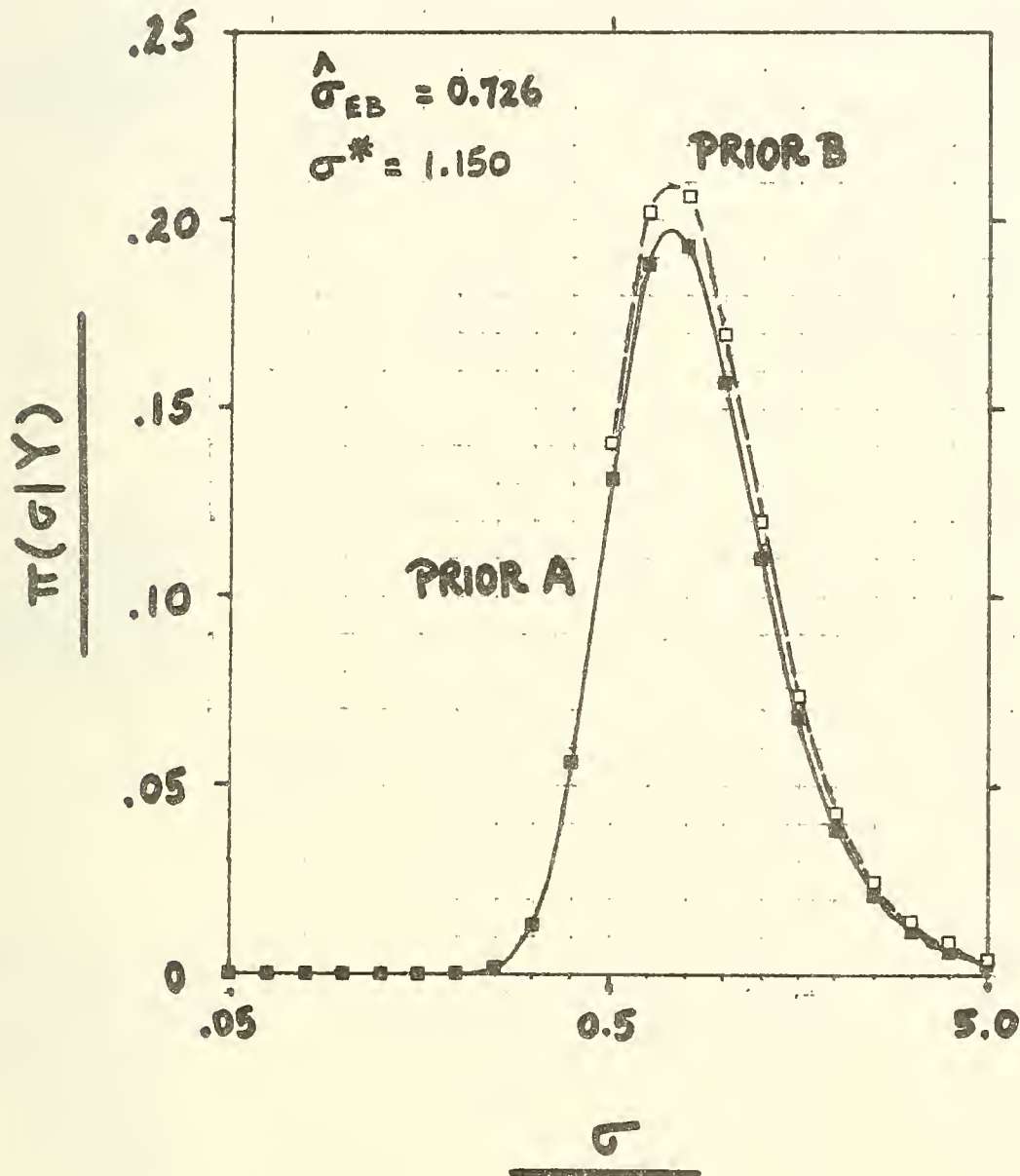
relative potency hypothesis. In the skin tumor initiation experiments, for example, roofing tar emission extracts were less potent than coke oven emission extracts. In the viral transformation studies, roofing tar emission extracts were more potent than coke oven emissions extracts.

Figure 3 shows the posterior densities $\pi(\sigma|Y)$ corresponding to this 3×3 experimental data matrix. The results for both prior distributions A and B, described in Section 4 above, are shown. We continue to assume a diffuse prior on β . In contrast to the 2×2 case, the posterior distribution of σ is considerably less sensitive to the prior distribution that is assumed. The likelihood function is now more concentrated around $\hat{\sigma}_{EB} = 0.726$. In the range $\sigma < 0.2$, the posterior density of σ is virtually zero.

These findings are reflected in Figure 4. Like Figure 1, this figure shows the means and standard errors of the original data on a logarithmic scale. Superimposed on these data are the posterior mean values of $X\beta$, derived from Prior A. Within each species, consecutive pairs of these posterior mean data points have been connected by dashed lines. Since the data points for each emission are equally spaced along the horizontal, and since the three logarithmic vertical axes are drawn to the same scale, the underlying constant relative potency hypothesis requires that the dashed lines connecting each pair of estimates be parallel. As Figure 4 shows, the underlying model predictions $X\beta^*$ in effect strike a balance between the contradictory elements in the original data.

Table 4 shows selected statistics of the posterior distributions of $X\beta$ and θ for the human lung cancer slopes, based on Priors A and B. Also shown are the empirical Bayes estimates corresponding to (3.16), where empirical Bayes estimate 1 uses $\hat{\sigma}_{EB}$ and empirical Bayes estimate 2 substitutes $\hat{\sigma}_{RSS}$

FIGURE 3 : Posterior Densities of $\log \sigma$
3 x 9 EXPERIMENTAL DATA MATRIX

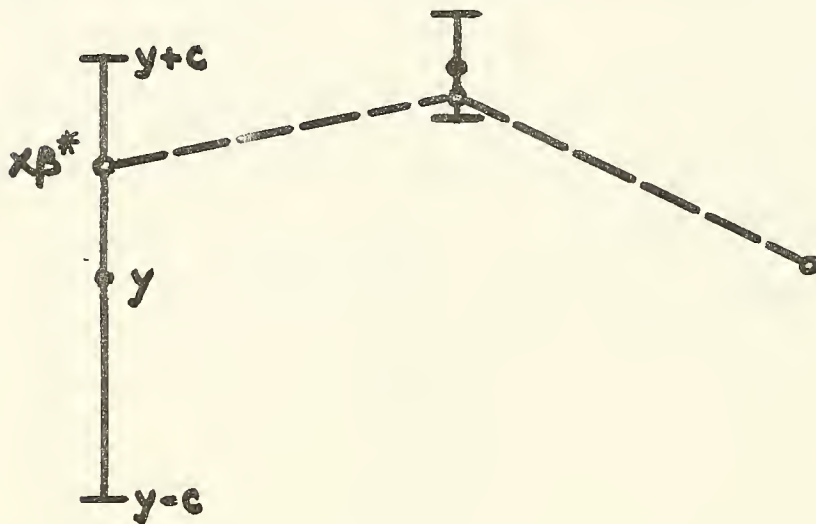


Prior A: $\log \sigma$ uniform on $0.05 \leq \sigma \leq 5.0$

Prior B: $\log \sigma$ uniform on $0.5 \leq \sigma \leq 5.0$

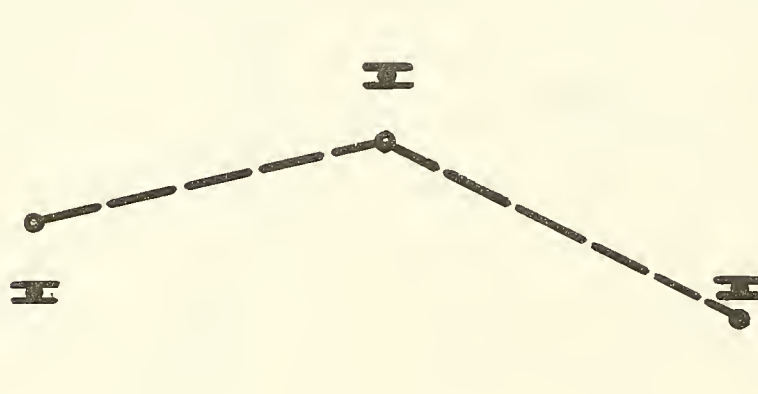
FIGURE 4
3 x 3 EXPERIMENTAL DATA MATRIX

incr. in rel. risk per 10^{-4} $\mu\text{g}/\text{m}^3$
extractable organics $\times$ years



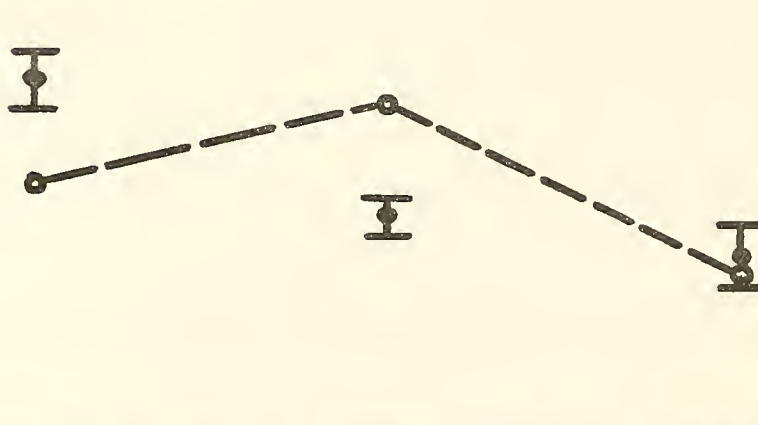
LUNG
CANCER
(MAN)

papillomas/mouse
per mg extract



SKIN
TUMOR
INITIATION
(MICE)

transformations/
 2×10^6 cells per $\mu\text{g}/\text{ml}$



ENHANCEMENT
OF VIRAL
TRANSFORMATION
(HAMSTER
EMBRYO
CELLS)

ROOFING
TAR

COKE
OVEN

AUTO-
MOTIVE

TABLE 4.

Bayes and Empirical Bayes Estimates of Log Slopes
For Lung Cancer Risk in Man
3 x 3 Data Matrix

Environmental Emission	Mean	Stand. Dev.	Poster. Mean $X\beta$	Lower Tail	Upper Tail
Roofing Tar					
Original Data	0.495	1.415			
θY (Prior B)	0.818	1.058	0.945	0.014	0.010
θY (Prior A)	0.832	1.036	0.952	0.015	0.010
θY (E.Bayes 1)	0.884	0.960	0.987		
θY (E.Bayes 2)	0.861	0.995	0.972		
Coke Oven					
Original Data	1.482	0.341			
θY (Prior B)	1.463	0.337	1.336	0.010	0.010
θY (Prior A)	1.462	0.336	1.341	0.010	0.010
θY (E.Bayes 1)	1.459	0.336	1.356		
θY (E.Bayes 2)	1.460	0.336	1.348		
Diesel Engine					
θY (Prior B)	0.434	1.875	0.434	0.017	0.015
θY (Prior A)	0.442	1.818	0.442	0.017	0.016
θY (E.Bayes 1)	0.466	1.217	0.466		
θY (E.Bayes 2)	0.454	1.300	0.455		

$\theta|Y$ (Empirical Bayes 1) assumes prior $\pi(\sigma)$ concentrated at $\hat{\sigma}_{EB}$
= 0.726, and diffuse prior on β .

$\theta|Y$ (Empirical Bayes 2) assumes prior $\pi(\sigma)$ concentrated at $\hat{\sigma}_{RSS}$
= 0.782, and diffuse prior on β .

for $\hat{\sigma}_{EB}$.

In comparison to the results for the 2x2 case (Table 2), the standard deviations of the posterior distributions for the roofing tar log slope were considerably less sensitive to the estimation method used. For both roofing tar and coke oven emissions, the posterior mean values of θ now deviate from the corresponding posterior mean values of $X\beta$.

In the diesel engine case, however, the contrast between the Bayes and empirical Bayes estimates is more striking. Because there were no original epidemiological data in this case, the posterior precision of θ depends solely on our assumptions about the hyperparameter σ . Whereas $\hat{\sigma}_{EB} = 0.726$ and $\hat{\sigma}_{RSS} = 0.782$ for these data, the Bayes estimates are $\sigma^* = 1.150$, based on Prior A, and $\sigma^* = 1.189$, based on Prior B.

The scientist who voices skepticism at the close fit of the data in Figure 1 has, it appears, been vindicated. If we take σ to be its maximum likelihood estimate $\hat{\sigma}_{EB} = 0.726$, then extrapolations between species or environmental agents, we conclude, will be accurate only to a multiplicative factor of 2 with 68 percent probability and only to a multiplicative factor of 4 with 95 percent probability. If we take σ to be the Bayes estimate $\sigma^* = 1.150$ (based on Prior A), then such extrapolations, we conclude, can be accurate only to a multiplicative factor of $\exp(1.15) \approx 3$ with 68 percent probability and only to a multiplicative factor of $\exp(2 \times 1.15) \approx 10$ with 95 percent probability.

We are now in a position to contrast our statistical approach with others in the literature. Our equation (3.2) partitions the sources of variation among experiments into several components. We thus follow Cochran's (1980) suggestion that "the summary of a series of experiments calls mainly for experience in the analysis of variance." In our decomposition of these

sources of variation, however, we do not assume that the true values of the log slopes exactly obey an additive model. We assume only that the true values lie within $\pm\sigma$ of such a model with probability 0.68. Further, when confronted with the problem of estimating the slope for a particular experiment, we have no belief that the slope in question is at all unusual in its deviation from the underlying equal relative potency model. For us, the set of all such deviations is exchangeable. The distinction between the Bayes and the empirical Bayes approaches depends on our willingness either to assign a prior distribution to σ (then integrating with respect to σ) or to use a point estimate for σ as if it were known. We prefer the full Bayesian procedure, especially when there are relatively few experiments, since the uncertainty in σ is real and should contribute to our uncertainty about θ . This point is illustrated by the Bayes and empirical Bayes standard deviations of θ for diesel engine emissions in Table 4. On the other hand, when many experiments are combined, we expect that the choice of prior distribution and the choice between Bayes and empirical Bayes estimates will be less important. (See, e.g., Tiao and Zellner (1964).)

One procedure suggested by Lindley and Smith (1972) and Smith (1973a), which they describe as "modal Bayesian," amounts to the use of $\hat{\sigma}_{MLE}$ in our formula (3.16) for $\hat{C}_{EB}$. We would describe this approach as yet another version of empirical Bayes. Smith (1973a) shows that if σ is known, then the Bayesian confidence intervals will be shorter than the classical confidence intervals for θ . Our allowing for uncertainty in σ will tend to lengthen the confidence intervals, but they will still be shorter than the classical intervals for θ_i based solely on the sampling errors c_i from each experiment.

The fully Bayesian analysis of Smith (1973a) differs from ours in several

respects. Since he concentrates on the two-way table with no replications, the components of error that we call C and $\sigma^2 I$ are combined in that paper as if C were equal to zero. Moreover, we explicitly calculate the posterior distribution of σ in order to: (a) show the range of uncertainty remaining for σ ; (b) calculate the value $\hat{\sigma}_{EB}$ for use in our empirical Bayes procedure; and (c) determine $\sigma^{*2} = E[\sigma^2|Y]$, which can be interpreted as a Bayesian risk of interspecies extrapolation if loss is proportional to $\delta^2 = (\theta - X\beta)^2$. The statistic σ^* will also play a critical role in the diagnostic procedure of the next section.

But there are still two serious problems with our analysis of the data of Table 3. First, we note that the more precise data come from non-human experiments. One may legitimately protest that we have merely learned how accurately we can extrapolate from mouse skin to hamster embryo cells. At the very least, some test of the assumption of exchangeable extrapolation errors seems appropriate. Ideally, we should include the results of more precise human experiments in our analysis.

Second, we have so far said nothing about the choice of experiments to be included in the analysis. Harris (1981) selected these laboratory bioassays because they were considered to be valuable quantitative measures of carcinogenicity, and because tests of several related emissions were performed in the same laboratory. The U.S. Environmental Protection Agency had chosen these specific emissions as part of its diesel emission research program (Huisinigh et al., 1979). Although we have presented only a few experiments initially for expository purposes, it is hardly clear what would happen if we were to include many more experiments. What is more, there is no obvious means of deciding which experiments are most appropriate to include.

6. SELECTING AND REJECTING EXPERIMENTS

6.1 The 5x9 Case

Table 5 further augments the experimental data in Table 3. In addition to lung cancer epidemiological studies in man, skin tumor initiation experiments in mice, and viral transformation studies in SHE cells, we have included mutagenesis experiments in L5178Y mouse lymphoma cells performed under two types of conditions (Mitchell et al., 1979). In the row denoted Mutagenesis-MA, no metabolic activator was added. In the row denoted Mutagenesis+MA, metabolic activator was included in the experimental preparation. Thus, both direct and indirect mutagenicity were measured.

In addition to the three emissions given in Table 3, we have included three other diesel engine emission samples; a sample of particulate emissions from a gasoline-powered automobile engine; the polyaromatic hydrocarbon benzo(a)pyrene; and cigarette smoke condensate from the Kentucky 1A1 experimental cigarette, which was designed to be typical of cigarettes smoked during the 1950s. The diesel engine extract appearing in Table 3 has been relabeled Diesel I, while the remaining diesel emission samples have been numbered from II to IV. Diesel emissions II and III were, like Diesel I, obtained from light duty diesel engines. Diesel emission IV was obtained from a heavy duty diesel engine. The conditions of collection of these samples are described in Huisinigh et al. (1979). With the exception of the results for cigarette smoke condensate, all dose-response slopes and the standard errors are taken from Harris (1981).

Although Harris (1981) did not report the corresponding dose-response slopes for cigarette smoke condensate, experiments on this agent were reported in the source studies (Casto et al., 1979; Mitchell et al., 1979; Nesnow et al., 1979). We were therefore able to estimate these slopes by the same

TABLE 5.
5 x 9 Experimental Data Matrix

	Roofing Tar	Coke Oven	Diesel Engine I	Diesel Engine II	Diesel Engine III	Diesel Engine IV	Gasol. Engine	Benzo(a) pyrene	Cig. Smoke*
Lung Cancer (Man)	1.64 1.41 0.49	4.40 0.34 1.48							0.03 0.15 -3.46
Skin Tumor Initiation (Sencar mice)	0.54 0.04 -0.63	2.10 0.04 0.74	0.53 0.04 -0.64	0.16 0.22 -1.86		0.01 0.82 -4.51	0.03 0.26 -3.61	85.28 0.03 4.45	0.00 1.30 -5.88
Enhancement of Viral Transform. SHE Cells)	2.07 0.18 0.73	0.86 0.10 -0.15	0.65 0.15 -0.44	0.07 0.33 -2.70	0.13 0.18 -2.06	0.04 0.59 -3.24	0.20 0.12 -1.59	540.00 0.04 6.29	0.58 0.08 -0.54
Mutagenesis -MA (Mouse Lymphoma Cells)**	0.31 0.39 -1.17	0.73 0.21 -0.32	1.66 0.31 0.51	0.27 0.43 -1.31	2.55 0.16 0.93	0.16 0.24 -1.86	0.35 0.11 -1.06		0.59 0.23 -0.53
Mutagenesis +MA (Mouse Lymphoma Cells)**	9.56 0.16 2.26	9.96 0.07 2.30	1.87 0.26 0.63	0.76 0.14 -0.27	1.01 0.20 0.01	0.05 0.43 -3.02	0.99 0.10 -0.01		0.45 0.13 -0.79

*Whole cigarette smoke condensate substituted for organic solvent extracts.

**Average mutant colonies/10⁶ survivors per µg/ml extract.

MA = metabolic activator.

Units for other rows as in Tables 1 and 3.

maximum likelihood methods used by Harris (1981). For the human lung cancer results, we applied Harris's estimation procedure to the U.S. veterans study data for men aged 35 to 84 observed during 1954 to 1962 (Kahn, 1966, Appendix Tables A to D). If $h(d,t)$ is the incidence of lung cancer among men of age t with accumulated dose d , we obtained a maximum likelihood estimate of the parameter ξ for the relative risk model

$$h(t,d) = h(t,0) (1 + \xi d).$$

(The estimation algorithm is described in DuMouchel (1981).) With accumulated dose measured in cigarettes per day $\times$ years, our maximum likelihood estimate for ξ was 1.085×10^{-2} units of incremental relative risk per cigarettes/day $\times$ years (standard error, 0.103×10^{-2}). This estimate was then converted into units of incremental relative risk per $10^{-4} \mu\text{g}/\text{m}$ cigarette smoke condensate $\times$ years under the assumption that the typical cigarette smoked by a subject delivered 38 ± 2 mg cigarette smoke condensate, and that the total daily delivery of condensate was diluted in a total daily ventilation of $11 \pm 2 \text{ m}^3$. (We used the methods described by Harris to incorporate the uncertainty about these dosage conversion units into the slope and coefficient of variation reported in Table 5.)

Except for the last two columns, the additional experiments in Table 5 were performed, as above, on the dichloromethane extracts of the various emissions. For the benzo(a)pyrene results, this agent was applied in concentrated form as a positive control in some experiments. For the last column, whole smoke condensate was used. The resulting dose-response units, we note, are still compatible with the constant relative potency model. For any two pairs (k,l) and (k',l') , the quantity $\theta_{kl} - \theta_{k'l} - \theta_{kl'} + \theta_{k'l'}$

remains dimensionless.

Table 6 reports the Bayes estimates of the log slopes for human lung cancer risk. The first two columns (denoted 2x2 and 3x3) summarize the results of the previous two sections. The third column provides the corresponding results for the 5x9 experimental data matrix in Table 5. The right-most column shows the original data. The remaining columns will be described momentarily. Only the results for roofing tar, coke oven, and diesel engine I emissions are given. The original slope for the human lung cancer risk from cigarette smoke was so precise that its posterior density did not change substantially. Hence, it is not reported.

The empirical Bayes estimate of σ for the 5x9 data matrix was 1.041. Because the posterior density $\pi(\sigma|Y)$ was highly concentrated around $\hat{\sigma}_{EB}$, the corresponding value of σ^* was nearly equal to $\hat{\sigma}_{EB}$. In comparison to the 3x3 case, the standard deviation of the posterior density of the roofing tar log slope has increased slightly. By contrast, the corresponding standard deviation for diesel engine I has declined. These results reflect a balance between two sources of uncertainty about θ . On the one hand, a large posterior value of σ implies uncertainty in the deviation δ . On the other hand, the larger number of experiments permits us to estimate $X\beta$ more precisely.

Figure 5 depicts the deviations in these data from the underlying constant relative potency model. For each of the five species, the Figure shows the posterior mean values of the residuals $\delta = \theta - X\beta$ for each emission. When there are no data y for a particular species-emission pair, the posterior mean of δ is necessarily zero. Such cases are therefore omitted from the Figure.

The posterior mean residuals for cigarette smoke, Figure 5 shows, are in

TABLE 6.

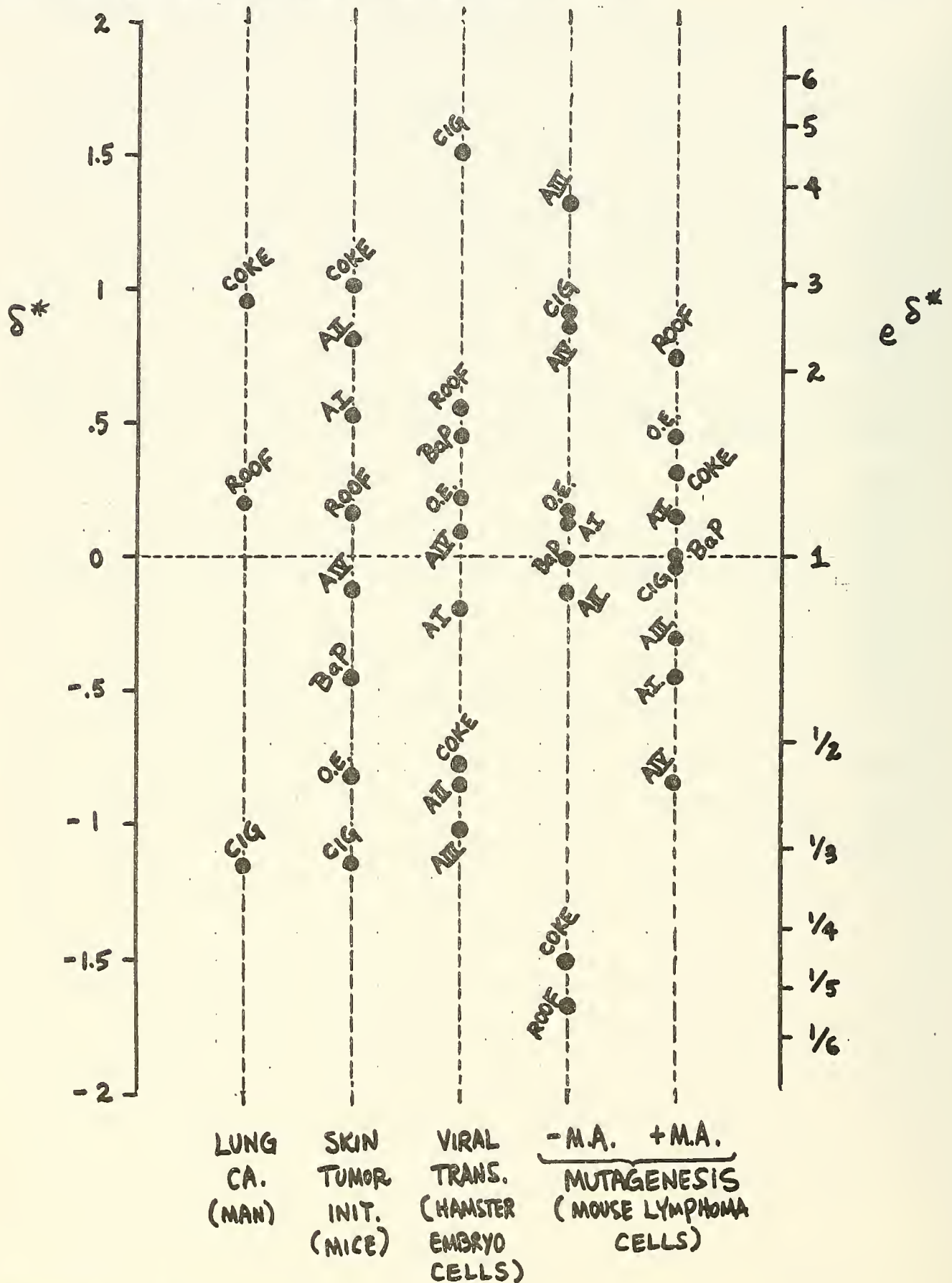
Bayes Estimates of Log Slopes For Lung Cancer Risk in Man
Based on Alternative Data Matrices

	2x2	3x3	5x9	4x9	4x8	3x8	3x7	Data
$\hat{\sigma}_{EB}$	0.0	0.726	1.041	0.872	0.674	0.389	0.316	
σ^*	1.043	1.150	1.080	0.933	0.730	0.480	0.395	
Roofing Tar								
θ^*	0.229	0.832	0.123	0.306	0.959	1.526	0.497	0.495
c^*	0.788	1.036	1.108	0.957	0.915	0.742	1.414	1.415
Coke Oven								
θ^*	1.497	1.462	1.375	1.366	1.455	1.422		1.482
c^*	0.334	0.336	0.335	0.334	0.335	0.334		0.341
Diesel Engine I								
θ^*		0.442	-0.458	-0.706	0.207	0.330	-0.836	
c^*		1.818	1.451	1.304	1.160	0.867	1.582	

Prior A for $\pi(\sigma)$ and diffuse prior for β assumed for all calculations.

$$\sigma^* = E[\sigma^2|Y]^{1/2}.$$

FIGURE 5: POSTERIOR MEAN RESIDUALS $\delta^* = \theta^* - X\beta^*$ DERIVED FROM 5×9 EXPERIMENTAL DATA MATRIX



four of five cases relatively large in absolute value. Cigarette smoke is apparently a weaker human lung carcinogen, a weaker mouse skin tumor initiator, a more potent transforming agent, and a more potent direct mutagen than would be predicted in this case from the constant relative potency model. Further, the range of the mean residuals is largest for the mutagenesis experiments in the absence of metabolic activator. Tests for indirect mutagenicity are apparently least compatible with the underlying model. By contrast, the residuals for mutagenesis with activator are more concentrated around the origin. A similar finding applies to the viral transformation results when cigarette smoke is eliminated.

6.2 A Diagnostic Procedure.

We seek a method to determine which subset of experiments is most relevant for predicting lung cancer risks in man. Two basic characteristics, we suggest, are critical to such a procedure.

First, the method ought to be sensitive to the underlying tradeoff between predictive bias and predictive precision. Suppose that we are interested in a particular θ_i for which there is little or no data (i.e., c_i is large). The inclusion of irrelevant experiments in the data matrix could result in a biased estimate of θ_i , the size of this bias being in the order of $\pm\sigma$. However, if we eliminated all but the most relevant experiments, the remaining experiments could contribute little if anything to the accuracy of our estimate of θ_i , as measured by c_i^* . This difficulty applies especially to the case where we have no original data y_i on θ_i (e.g., the human lung cancer risks for diesel engine emissions in Table 5). If we eliminated every conceivably irrelevant experiment, then we would end up with exactly what we had at the start—no information on θ_i at all.

Second, we should not eliminate experiments individually. One could exclude those species-emission pairs with large posterior mean values of δ^2 . Since these interactions are what determines the relatedness of the various species and emissions, such a procedure would defeat the purpose of our analysis. It would be more appropriate to assess whether a specific species or a specific environmental agent is more or less relevant to the others. We therefore adopt a cross-validation procedure based on the elimination of entire rows or columns from the matrix of experiments.

Let Y_{k-} be the vector of log slopes formed by exclusion of all experiments involving species k . For each species k , we evaluate the posterior density $\pi(\sigma|Y_{k-})$, and denote $\sigma_{k-}^* = E[\sigma^2|Y_{k-}]^{1/2}$. We shall say that species k is "less relevant" if $\sigma_{k-}^* < \sigma^*$, where $\sigma^* = E[\sigma^2|Y]^{1/2}$ as in (4.1). Analogous definitions apply to $Y_{-\ell}$ and $\sigma_{-\ell}^*$ for each environmental agent ℓ . The species or agent for which σ_{k-}^* or $\sigma_{-\ell}^*$ is lowest will be termed the "least relevant". The least relevant species or emission is the one whose elimination most improves the relevance of the remaining experiments to each other. In anticipation of Section 7, we note that the terms less relevant and least relevant are a posteriori concepts.

Given an initial set of experiments Y , a prior density $\pi(\sigma)$, and a particular θ_i of interest, we consider the following data analytic procedure.

(i) Calculate σ_{k-}^* and $\sigma_{-\ell}^*$ for each k and ℓ , and determine the least relevant species or emission.

(ii) Calculate the posterior distribution of θ_i before and after the least relevant species or emission is removed.

(iii) Eliminate the least relevant species or emission and repeat steps (i) and (ii) on the reduced set of experiments so long as: (a) there exists a

less relevant experiment; (b) the least relevant species or emission does not correspond to θ_i ; and (c) the elimination of the least relevant species or emission reduces σ_i^* , the posterior standard deviation of θ_i .

(iv) If conditions (a), (b), and (c), are not satisfied, the procedure terminates. The remaining experiments are considered most relevant for predicting θ_i .

We applied this procedure to the 5x9 data matrix in Table 5. Prior A on σ was assumed. We focused on predicting the human lung cancer slopes for roofing tar emissions and diesel I emissions.

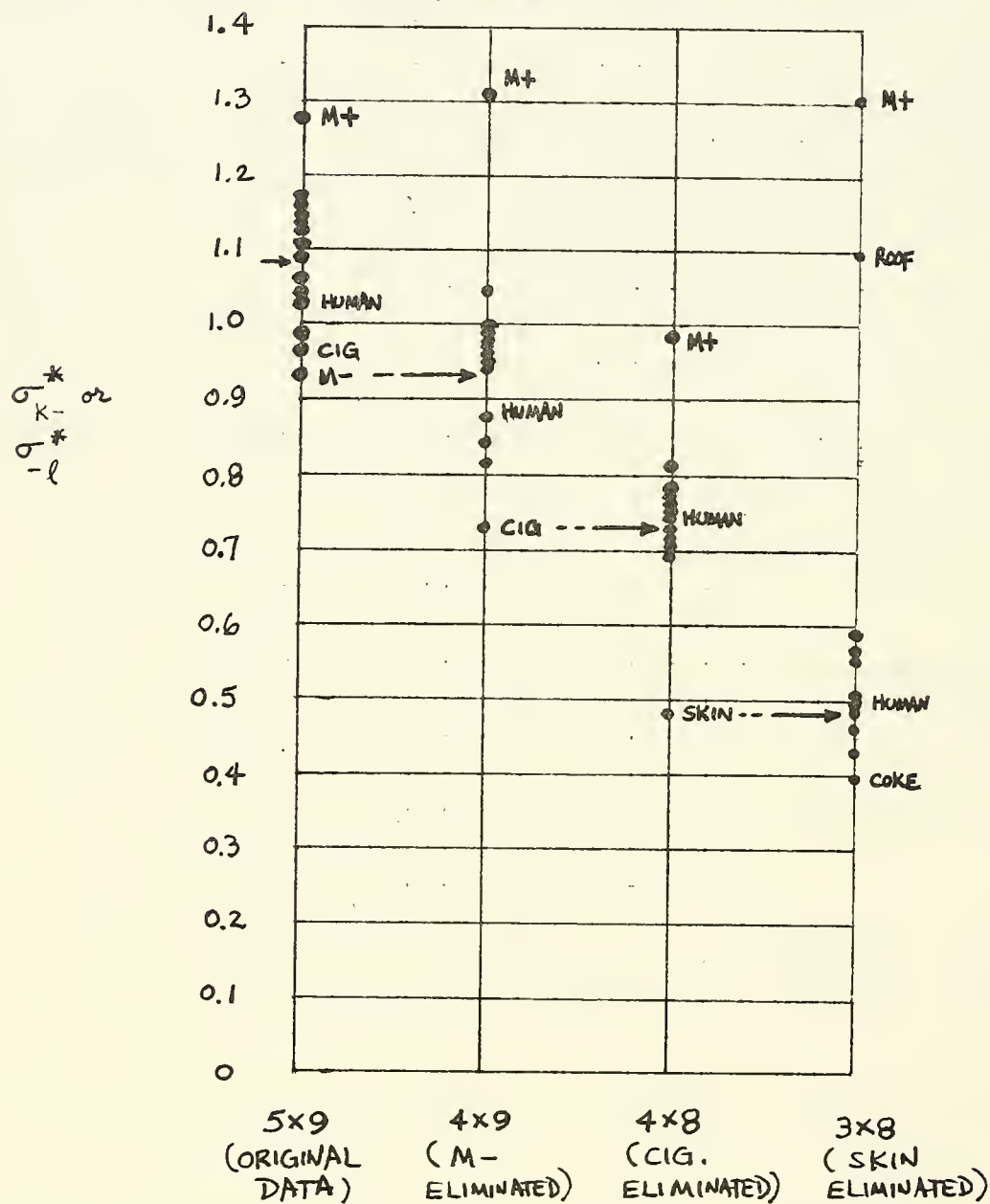
Steps (i), (ii), and (iii) were repeated four times. Figure 6 depicts the distribution of values of σ_{k-}^* and σ_{-l}^* for each iteration, where successive iterations are displayed from left to right. To assist interpretation, a few species and agents are specifically identified.

For the original 5x9 data matrix, $\sigma^* = 1.08$. Mutagenesis without metabolic activation was least relevant. Removal of this row resulted in a 4x9 matrix with a new $\sigma^* = 0.933$. Repeating this procedure, we found cigarette smoke to be least relevant. Removal of this column resulted in a 4x8 matrix with a new $\sigma^* = 0.730$. Again repeating this procedure, we found skin tumor initiation in mice to be least relevant. Removal of this row resulted in a 3x8 matrix with a new $\sigma^* = 0.480$. In the final iteration, coke oven emissions were found to be least relevant, with $\sigma_{-coke}^* = 0.395$. Elimination of coke oven emissions from the 3x8 matrix violated condition (c) in step (iii) above. Hence, the procedure was terminated and the 3x8 array was deemed most relevant.

The results of this procedure are summarized in those columns of Table 6 labelled 5x9, 4x9, 4x8, 3x8, and 3x7. As we successively eliminate mutagenesis without activator, cigarette smoke, and skin tumor initiation, the

FIGURE 6

Distribution of σ_{k-}^* and σ_{-l}^* for each iteration of the diagnostic procedure.



$$\sigma_{k-}^* = E[\sigma^2 | Y_{k-}]^{1/2} \quad \text{where } Y_{k-} \text{ is experimental data exclusive of species } k.$$

$$\sigma_{-l}^* = E[\sigma^2 | Y_{-l}]^{1/2} \quad \text{where } Y_{-l} \text{ is experimental data exclusive of agent } l.$$

The arrows point to the value of σ^* when all data are included.

posterior standard deviations of roofing tar and diesel I emissions decline. Elimination of coke oven emissions, the least relevant in the 3x8 array, resulted in a marked increase in the posterior standard deviations of the roofing tar and diesel emissions parameters. In the case of roofing tar, the estimates θ^* and c^* were almost identical to the original values of y and c .

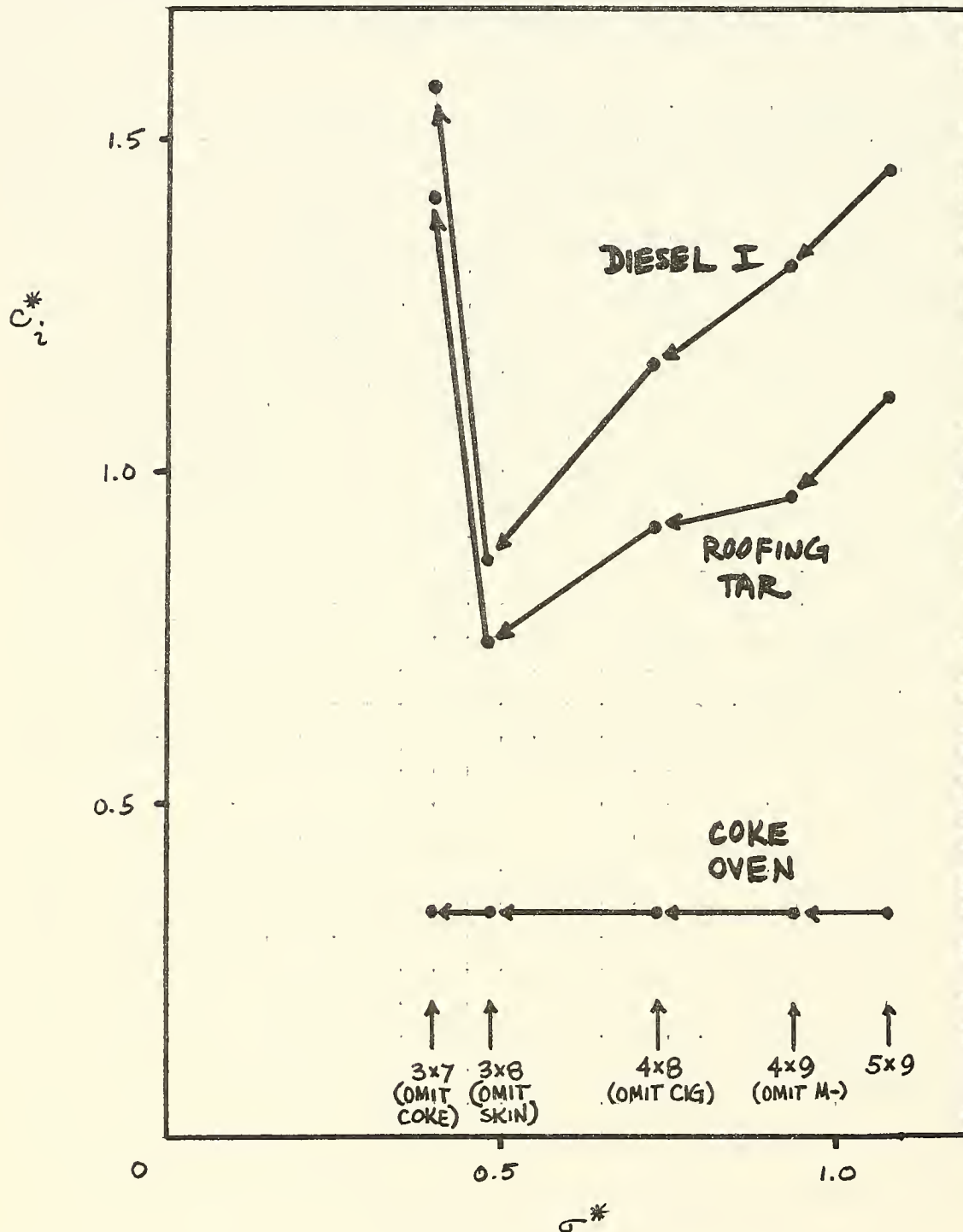
The resulting tradeoff between predictive bias (σ^*) and predictive efficiency (c_i^*) is depicted graphically in Figure 7. By successive elimination of least relevant experiments, we are able to reduce σ^* to 0.48. Any further reduction in σ^* is at the cost of a marked loss of precision.

Unless we are willing to specify a particular loss function, we cannot unequivocally conclude that the predictions resulting from the 3x8 matrix are most preferred. For many public health and environmental policy applications, however, a reduction in the extent of uncertainty about human risks is desired. To seek to eliminate less relevant experiments, so long as predictive efficiency is reduced, appears to be appropriate for such situations. (Our procedure depends somewhat upon the choice of prior distribution $\pi(\sigma)$, but when many experiments are involved, this dependence should be minimal.)

The human lung cancer experiments, we note from Figure 6, are less relevant (i.e. $\sigma_{human}^* < \sigma^*$) so long as cigarette smoke is included in the data matrix. This conclusion does not apply to the 4x8 or 3x8 arrays, with cigarette smoke removed. Cigarette smoke contains numerous carcinogenic and mutagenic compounds other than polyaromatic hydrocarbons, e.g., nitrosamines and various heterocyclics. The apparent deviations of the cigarette smoke data from the constant relative potency model may reflect these differences in chemical composition. Peto (1977) has similarly remarked that the mutagenic

FIGURE 7

Tradeoff between σ^* and c_i^* for each iteration of diagnostic procedure



potency of cigarette smoke appears to be greater than would be expected from its systemic carcinogenicity. Whatever its interpretation, our procedure leads us to eliminate the most precise human experiment. As a consequence, our estimate of σ is constructed primarily from the non-human data. We see no completely satisfactory response to this limitation other than to suggest, where possible, the inclusion of other precise human data.

Nevertheless, we find the results of our diagnostic procedure intriguing. Assays for indirect mutagenicity and tumor initiation have been excluded as less relevant. The remaining laboratory bioassays are designed to gauge an agent's interference with gene replication and cell differentiation. For the polycyclic aromatic-containing emissions remaining in the 3x8 table, these biological processes could be critical to human lung carcinogenesis.

We recognize that the above cross-validation procedure is purely data analytic. Adapting the methods in Efron and Morris (1973) to the current problem, we could replace our assumption that $V(\delta_i) = \sigma^2$ (for all i) with a more general specification. That is, we might assume $V(\delta_i) = \tau^2$ for some subset of experiments and then employ a joint prior distribution for (σ, τ) to derive the posterior distribution of θ . Unless the "suspicious subset" can be identified a priori, our present method appears to be much simpler in practice.

7. PERFECTLY REPLICATED, IMPERFECTLY REPLICATED, AND STRONGLY RELATED EXPERIMENTS.

7.1 Definitions.

In some situations, we may have additional prior information on the relationships between experiments. In Table 5, for example, it is not unreasonable to posit that diesel engine emissions I through IV ought to be more related to each other than to the remaining environmental agents. An

analogous assumption might apply if experimental data were available in two closely related species, or two strains of the same species, or even males and females of the same species.

We now investigate how one might take advantage of this form of prior information. For this purpose, we need to characterize precisely how experiments can be related a priori in terms of our statistical model. A rather different classification of degrees of relatedness is given by Smith (1973c).

Consider the results y_i of a particular experiment. Its mean θ_i can be separated into two components, $x_i\beta$ and δ_i , with $\theta_i = x_i\beta + \delta_i$. We shall call two experiments i and i' "perfectly replicated" if $x_i\beta = x_{i'}\beta$ and $\delta_i = \delta_{i'}$ a priori. Under this definition, the only source of variation in $(y_i - y_{i'})$ is the sampling error associated with each experimental observation.

We shall call two experiments i and i' "imperfectly replicated" if $x_i\beta = x_{i'}\beta$ a priori, but, conditional on σ , the hyperparameters δ_i and $\delta_{i'}$ are a priori independent $N(0, \sigma^2)$. If $x_i\beta$ is highly correlated a priori with $x_{i'}\beta$, we shall say that the corresponding experiments are "strongly related." The term "highly correlated" is temporarily left vague. Finally, all pairs of experiments that are neither perfectly replicated, imperfectly replicated, or strongly related will be defined as "weakly related."

The case of perfectly replicated experiments presents no special problems for the present paper. If $y_i \pm c_i$ and $y_{i'} \pm c_{i'}$ are the sufficient statistics for the two experiments, we merely replace them by the single statistic $y_{i''} \pm c_{i''}$, where

$$y_{i''} = (c_i^2 y_i + c_{i'}^2 y_{i'}) / (c_i^2 + c_{i'}^2),$$

$$c_{i''} = (c_{i'}^{-2} + c_{i''}^{-2})^{-1/2}.$$

We might apply this procedure if two different experiments were independently performed in the same species with the same agent but, say, in different laboratories. This case will not be considered further.

When two experiments are imperfectly replicated, we admit the possibility of independent deviations from the underlying regression model, as well as independent sampling errors. When the hyperparameter β has prior distribution $N(b, V)$, imperfect replication implies that (conditional on σ) the slopes θ_i and $\theta_{i'}$ have a priori identical prior means x_b , identical variances $xVx' + \sigma^2$, and correlation coefficient

$$(7.1) \quad xVx' / (xVx' + \sigma^2),$$

where $x_i = x_{i'} = x$ are row vectors.

When a diffuse prior is assumed for β , we must apply the definition of imperfect replication with care. If V is replaced by tV and $t \rightarrow \infty$, the a priori correlation coefficient (7.1) approaches unity. Since θ_i and $\theta_{i'}$ are normally distributed with identical means and variances, a correlation coefficient of 1 would imply that $\theta_i = \theta_{i'}$, a priori. But this would imply that $\theta_i = \theta_{i'}$, a posteriori, even if $(y_i - y_{i'}) / (c_i^2 + c_{i'}^2)^{1/2}$ is large. The assumption of a diffuse prior on β apparently reduces the notion of imperfect replication to that of perfect replication, even though δ_i and $\delta_{i'}$ are assumed to differ on the order of σ .

This difficulty appears to be related to the class of marginalization paradoxes discussed by Dawid, Stone, and Zidek (1973), which are known to occur sometimes when improper prior distributions are employed. We note from

formula (3.14) in Section 3.3, however, that when β takes on a diffuse prior, $E[\theta|Y, \sigma] = (I+CR)^{-1}Y$, where $R = \sigma^{-2}(I-X(X'X)^{-1}X')$. It is straightforward to show that for this formula, $E[\theta_i - \theta_{i'}|Y, \sigma]$ is zero only when $y_i = y_{i'}$, even if $x_i = x_{i'}$. The paradox that θ_i and $\theta_{i'}$ are perfectly correlated a priori is avoided so long as we use (3.14) to evaluate the posterior means and variances of θ .

The case where two experiments are strongly related is even more general. The correlation between $x_i\beta$ and $x_{i'}\beta$ is a priori

$$(7.2) \quad r = x_i'Vx_{i'} / [(x_i'Vx_i)(x_{i'}'Vx_{i'})]^{1/2},$$

while (conditional on σ) the correlation between θ_i and $\theta_{i'}$ is a priori

$$(7.3) \quad x_i'Vx_{i'} / [(x_i'Vx_i + \sigma^2)(x_{i'}'Vx_{i'} + \sigma^2)]^{1/2},$$

which reduces to (7.1) when $x_i = x_{i'}$ (i.e., imperfect replication). Note that the correlation between θ_i and $\theta_{i'}$ in (7.3) is always less than r in absolute value.

We do not wish to draw a sharp boundary between the terms strongly related and weakly related. If the correlation r in (7.2) exceeds 0.9 in absolute value, we would certainly call the corresponding experiments strongly related. If $|r| < 0.7$, or $r^2 < 0.5$ (i.e., the "between $x\beta$ " variance falls below the "within $x\beta$ " variance), then we would use the term weakly related. When $x_i \neq x_{i'}$ a priori and a diffuse prior is assumed for β , i.e., the assumptions used in Sections 4, 5, and 6, we regard the corresponding experiments as weakly related.

7.2 Informative Priors on β For the 3x8 Case.

Figure 8 diagrams a 3x8 array of experiments similar to that derived from the diagnostic procedure in Section 6. As a result of our elimination of the skin tumor data, the remaining single experiment on benzo(a)pyrene could not affect the posterior distributions of the other slopes. Hence, it was removed. In its place, however, we have added the results of an epidemiological study of men exposed in their occupations to a fifth type of diesel engine, a heavy duty diesel different from the other diesels. This additional slope was taken from Harris's (1981) analysis of lung cancer incidence among London Transport Authority diesel bus workers. (The slope estimate was originally reported in units of incremental relative risk per $\mu\text{g}/\text{m}^3$ particulates x years. It was converted to units of incremental relative risk per $10^{-4} \mu\text{g}/\text{m}^3$ extractable organics x years under the assumption that the dichloromethane extractable fraction constitutes 18 percent of particulates by weight.) No bioassay studies of the latter type of diesel engine emission were available.

We wish to introduce our prior knowledge that the experiments in diesel columns I through V are more related to each other than to the remaining experiments. Just what degree of interrelatedness should be assumed is not clear. It is implausible that the diesel experiments in each row should be perfectly replicated. To assume that they are imperfectly replicated requires, in effect, that the diesel emissions are virtually identical in composition, but that different diesel engine samples, through variations in engine design or operating conditions, may result in idiosyncratic effects in some species. If we were interested primarily in the possible risks of exposure to light duty diesel emissions, the assumption that all light duty and heavy duty diesel experiments were imperfect replicates could be too

	ROOF	COKE	DIESEL I	DIESEL II	DIESEL III	DIESEL IV	DIESEL V	GASOLINE POWERED
LUNG CANCER (man)	X	X					X	
Viral Transf.	X	X	X	X	X	X		X
Mutagenesis + MA	X	X	X	X	X	X		X
$l =$	1	2	3	4	5	6	7	8

FIGURE 8

3x8 DATA MATRIX

"X" means experimental data available

restrictive. The assumption of strongly related experiments may therefore be more useful.

Of course we could fall back on the assumptions of unequal x 's and a diffuse prior on β . In that case, however, there is no point in including the observed slope for diesel V. Its value of y would always be perfectly fit to the underlying regression model by the additional column effect that must be estimated, and it would not contribute toward estimation of the other θ 's. Only a proper prior on β would reflect our belief that the data on the human occupational exposure to diesel engine V are relevant to the estimation of lung cancer from exposure to the other diesel engines.

We now examine in detail the choice of a proper prior for β . Return to equation (3.2) in Section 3.1, that is,

$$y_{k\ell} = \mu + \alpha_k + \gamma_\ell + \delta_{k\ell} + \varepsilon_{k\ell},$$

where $k=1,2,3$ correspond to the three species and $\ell=1,\dots,8$ correspond to the eight environmental agents in Figure 8. The hyperparameters $\{\gamma_3, \dots, \gamma_7\}$ refer specifically to the various diesel effects. A natural model for the relationship between these hyperparameters is

$$(7.4) \quad \gamma_\ell = \gamma_0 + \eta_\ell, \quad \ell=3, \dots, 7,$$

where γ_0 is a component common to all diesel engines and $\{\eta_3, \dots, \eta_7\}$ represent deviations of each diesel from the common component. Each η_ℓ has prior mean zero and is independent of the other η_ℓ and of γ_0 .

Now denote the prior variance of γ_0 by v_0 and the prior variance of η_ℓ by v_η . If $v_0 = 0$ a priori, then the hyperparameters $\{\gamma_\ell\}$ are

uncorrelated. The quantities $\{\mu + \alpha_k + \gamma_\ell\}$, i.e., the quantities $\{x_i\beta\}$, will be correlated only through the common hyperparameters $\{\mu, \alpha_k\}$. Provided that v_η is not small compared to the prior variances of $\{\mu + \alpha_k\}$, the diesel columns are weakly related.

On the other hand, if $v_\eta = 0$ a priori, then the hyperparameters $\{\gamma_\ell\}$ are perfectly correlated, that is, the diesel columns are imperfect replicates. Finally, if v_0 is large and $v_\eta > 0$ is small, then the diesel columns are strongly related.

Since there are $K+L+1 = 3+8+1 = 12$ components in the hyperparameter set $\{\mu, \alpha_k, \gamma_\ell\}$, we must specify a 12×1 vector b of prior means as well as a 12×12 prior covariance matrix V . This covariance matrix contains a 5×5 submatrix of covariances among the diesel column effects $\{\gamma_3, \dots, \gamma_7\}$. We now make the following numerical assumptions.

(i) For every component of the prior mean of $\{\mu, \alpha_k, \gamma_\ell\}$, we choose $b = 0$.

(ii) For all (j, j') that are not part of the 5×5 submatrix of diesel column effects, we choose $V_{jj'} = 100I_{jj'}$.

(iii) To exemplify weakly related experiments, we choose $v_0 = 0$ and $v_\eta = 100$. In this case, $V = 100I$, where I is the 12×12 identity matrix, and for any pair of diesel experiments in the same species, the correlation r , defined in (7.2), is 0.67.

(iv) To exemplify imperfectly replicated experiments, we replace (iii) with the assumption that $v_0 = 100$ and $v_\eta = 0$. In this case, the number of column effects L is reduced to 4, and the hyperparameter set $\{\mu, \alpha_k, \gamma_\ell\}$ could be reduced to only 8 components. The computations are equivalent to a 3×4 analysis, where the design matrix X has several identical rows for each diesel engine, and $V = 100I$, where I is the 8×8 identity matrix.

(v) To exemplify strongly related experiments, we replace (ii) and (iii) with the assumption that $v_0 = 99$ and $v_\eta = 1$. For any pair of diesel experiments in the same species, the correlation r , defined in (7.2), is 0.997. The prior covariance matrix V takes the form

$$V = \begin{array}{c|c|c} \begin{array}{ccccc} 100 & 0 & . & . & 0 \\ 0 & 100 & . & . & . \\ . & . & . & . & 0 \\ 0 & . & . & 100 & . \end{array} & \begin{array}{c} \\ \\ 0 \\ \\ \end{array} & \begin{array}{ccccc} 100 & 99 & . & . & 99 \\ 99 & 100 & . & . & . \\ . & . & . & . & 99 \\ 99 & . & . & 99 & 100 \end{array} \\ \hline \begin{array}{c} \\ \\ 0 \\ \\ \end{array} & \begin{array}{c} \\ \\ \\ \\ \end{array} & 100 \end{array}$$

where the upper left submatrix is the covariance of $\{\mu, \alpha_k, \gamma_1, \gamma_2\}$, the submatrix in the center is the covariance of $\{\gamma_3, \dots, \gamma_7\}$, and the lower right element is the variance of γ_8 .

We chose a value of 100 for the prior variance of the hyperparameters $\{\mu, \alpha_k, \gamma_\ell\}$ to reflect our near complete state of ignorance about these effects. (It is possible, as Smith(1973a) shows in a somewhat simpler situation, to let these variances go to infinity and still retain our concept of strongly related experiments for fixed v_η . Our choice of a large finite value for v_0 is more convenient here.) Under this prior assumption, every log slope θ_i has variance $300 + \sigma^2$, and therefore we are uncertain about the magnitude of each slope to a multiplicative factor of about $\exp(20) \approx 5 \times 10^8$. Since these variances are so large, the somewhat arbitrary choice of $b=0$ does not materially affect the calculations, since all values of y_i are within

± 5 of $X_b = 0$. Our choice of $v_\eta = 1$ in (v) reflects our sense of the likely magnitude of the difference between any two diesel log slopes in the same species (e.g., $\theta_{k3} - \theta_{k7}$). The a priori standard deviation of this difference is $\sqrt{2(v_\eta + \sigma^2)}$ conditionally on σ . Even if $\sigma = 0$, our choice of $v_\eta = 1$ means the ratio of the two slopes is $\exp(z\sqrt{2})$, where z is $N(0,1)$. (Note that it is essential to choose $b_1 = b_2 = \dots = b_{11}$ for this to be true.) Since $\exp(z\sqrt{2}) \approx 10$ for $z = 1.6$, our prior choice of $v_\eta = 1$ allows for more than a 10 percent chance that one of the two diesels is 10 times as potent as the other. No greater uncertainty seems justified. At the other extreme, to assume that $v_\eta \ll \sigma$ would be nearly equivalent to imperfect replication.

Table 7 shows the resulting estimates of σ and of the θ 's for roofing tar, coke oven, diesel I and diesel V emissions. Since the assumption that the diesel columns are weakly related approximates complete ignorance about β , the results in that column are quite close to those obtained in the diffuse prior case in Table 6 (see the column labelled 3x8). The log slope for the diesel V epidemiological study contributes to the estimation of σ and to the remaining θ 's only through the vague prior on β . Its own value of y is nearly perfectly fit to the underlying model.

The assumption of imperfect replication, however, results in a dramatic increase in the estimate of σ . Any variation in the diesel θ 's that is in fact due to the γ 's is forced to be fitted to the corresponding δ 's. Hence, the posterior variance of δ increases. Although this prior assumption reduces the posterior standard deviation for Diesel V, the precision of the other estimates deteriorates. The assumption of imperfect replication, we conclude, is inappropriate in the present illustration.

The assumption of strongly related experiments does not have this

TABLE 7.

Bayes Estimates of Log Slopes For Lung Cancer Risk in Man
 Alternative Assumptions on the Relation Between Diesel Columns
 3 x 8 Data Matrix

(ν_0, ν_η)	Weakly Related (0,100)	Imperfect Replicates (100,0)	Strongly Related (99,1)	Original Data
$\hat{\sigma}_{EB}$	0.388	1.100	0.416	
σ^*	0.478	1.221	0.515	
Roofing Tar				
θ^*	1.533	1.393	1.724	0.495
c^*	0.739	1.124	0.743	1.415
Coke Oven				
θ^*	1.424	1.504	1.495	1.482
c^*	0.333	0.336	0.330	0.341
Diesel Engine I				
θ^*	0.339	-0.401	0.442	
c^*	0.862	1.594	0.868	
Diesel Engine V				
θ^*	1.877	0.443	0.241	1.921
c^*	1.497	1.165	1.029	1.512

limitation. Our conservative use of prior information about the relatedness of the diesel experiments does not increase σ^* much beyond 0.5. The posterior standard deviations of the roofing tar, coke oven and diesel I emissions are increased only slightly in comparison to the first column of the table. But the precision of the estimate for diesel V is improved. The incorporation of epidemiological data on occupational exposures to diesel engine V has led us to revise upward our estimate of the slope for diesel I. Moreover, the results of the viral transformation and mutagenesis experiments on diesels I through IV have led us to revise downward our estimate of the potency of diesel V. Rather than being more potent than roofing tar or coke oven emissions, as originally suggested by the data, diesel V is likely to be 3 or 4 times less potent, although not conclusively so.

8. DISCUSSION AND CONCLUSIONS

We have constructed a general framework for combining the results of diverse experiments when there is uncertainty about the relevance of some experiments to others. Within this framework, we have attacked the specific problem of assessing human cancer risks from heterogeneous toxicological and epidemiological data.

We distinguish between the conventional sampling error inherent in each experiment and a novel error of imperfect relevance among experiments. The latter type of error formalizes our notion of the credibility of interspecies and interagent extrapolations. We show how the available experimental data, in combination with the scientist's prior information on the credibility of such extrapolations, can be used to estimate the effects of various environmental agents in man and other species.

For a relatively simple example involving two species and two

environmental agents, we show how the scientist's prior information can override the data in predicting human cancer risks. When we add more experiments on a third agent and in a third species, the data are predominant. At the same time, a scientist's vague prior notion of the magnitude of these extrapolative errors is made more precise.

We then propose a data analytic method for selecting the most relevant subset among a multitude of experiments. The main idea behind this method is to determine which species or which environmental agent contributes most to our estimate of extrapolative error. Such species or agents are successively eliminated from the data base so long as the precision of a particular estimate, say, a particular human cancer risk, is improved.

We apply this diagnostic method to a relatively large 5×9 array, containing 36 observed dose-response slopes. This example demonstrates the tradeoff between prediction bias due to potentially irrelevant experiments and prediction efficiency resulting from combining diverse experiments. The analysis tentatively suggests that for a particular class of environmental emissions containing polycyclic aromatic hydrocarbons, the results of mammalian cell transformation experiments and mammalian mutagenesis experiments with metabolic activator are more relevant to human lung cancer risks than mammalian skin tumor initiation experiments or tests of indirect mutagenicity. Although this finding may not ultimately withstand scrutiny, it was derived, we stress, from our adoption of an attitude of exploratory analysis.

Finally, we demonstrate how prior information on the relationship between experiments can be incorporated into the analysis. This situation is likely to occur when experiments have been performed in the same species or in different strains of the same species, or when tests have been performed on multiple samples of the same environmental mixture.

The main limitation of the present analysis, we feel, is our inability to verify that the assumption of exchangeable errors of interspecies extrapolation applies to humans. The reader could legitimately object that we have merely assessed how well one can extrapolate from mouse skin to hamster embryo cells to mouse lymphoma cells. To confirm that the error of extrapolation is exchangeable among species, we need precise human carcinogenesis data.

Our analysis in Section 6 revealed that cigarette smoke is a more potent direct mutagen and a more potent transforming agent in cell culture than would be expected from its observed carcinogenic potency in man. When we excluded experiments involving cigarette smoke, the estimate of σ for the remaining data declined. We do not regard this finding as strong evidence against exchangeability of extrapolation errors across all species. However, the conclusion that the remaining data fit the underlying model more exactly, we acknowledge, is based primarily on the more precise non-human experimental data.

We should note, however, that the assumption of spherical errors (equation (3.3c)) is only a special case. The covariance matrix $\sigma^2 I$ for the extrapolative errors δ could be replaced by a more general matrix. For example, the deviations δ corresponding to experiments in a particular species or agent could have a variance that is different a priori from σ^2 . We have not explored these possibilities in this initial paper because the naive exchangeability assumption seems to be a reasonable starting point.

Moreover, we do not regard the basic normal data, normal prior structure of our model as particularly objectionable. Deviations from the underlying constant relative potency model may arise from biological processes that are non-gaussian. Since the normal distribution has smaller tails than other

likely candidates, any outliers from the underlying normal model will have a stronger contribution to the overall estimate of the hyperparameter σ . Our use of the normal model is thus more conservative in this respect. In any case, it gives analytical formulas that permit others to reproduce our results.

Nor do we take the underlying constant relative potency model to be an important limitation. Our framework could easily accommodate a constant additive potency model or, for that matter, any regression model of the form $\theta = X\beta + \epsilon$, where X is a known design matrix. The appeal of the constant relative potency concept is its avoidance of potentially complex or implausible conversions of dosage units between species.

Nor do we attach any special limitation to our apparent reliance on the slope of a linear dose-response relationship. Although we recognize that there is considerable support for such a dose-response model, it should be clear that alternative methods of summarizing the results of an experiment are possible. For example, the TD_{50} (dose at which 50 percent of subjects develop clinical toxicity) or the MTD (maximum tolerated dose) could be used for each human experiment, as in Freireich et al. (1966). For the non-human species, at least, the LD_{50} could be used, as in Meselson and Russell (1977). In fact, our model could be generalized to the multivariate case where each experiment is summarized by a vector of numbers. In this way, one could incorporate the effect of such additional factors as the effects of duration and fraction of exposure, or possible synergistic effects with other tumor initiators or promoters.

The model described in this paper appears to provide the theoretical underpinning for the other previous attempts to combine carcinogenesis experiments. Meselson and Russell's (1977) comparison of mutagenic potency in

Salmonella with carcinogenic potency in rodents constitutes a special 2xL case in our framework. In fact, the present methodology can be used to resolve the criticism that the favorable results of these authors were merely fortuitous.

Similarly, our approach resolves the difficulties encountered by Crouch and Wilson (1979,1980) in having to perform separate comparisons of carcinogenic potency in different pairs of species. It also satisfies these authors' desire for a systematic approach to the identification of potential exceptions to the underlying extrapolative model. Moreover, the use of informative priors on the hyperparameters of the model, as illustrated in Section 7, permits us to include multiple experiments on the same agent in the same species. We therefore avoid the problem, encountered by these authors, of deciding which of several experiments to incorporate in the analysis. By the use of informative priors, we could also incorporate information about the faulty design or execution of an experiment. Furthermore, in a multivariate generalization of our model, we could incorporate the incidences of tumors of different sites. This would avoid the additional difficulty, encountered by these authors, of deciding which of several endpoints to choose.

Finally, this paper considers only the estimation of carcinogenic potency. We do not discuss the use of these estimates, in combination with data on the extent of exposure to an environmental agent, to predict possible excess cancer incidence. Such an application entails additional but important uncertainties that are beyond the scope of this study.

9. ACKNOWLEDGEMENTS

Prof. DuMouchel's research was supported by National Science Foundation Grant No. MCS-80-05483. Prof. Harris's research was supported by Public Health Service Research Grant No. DA-02620 and by Research Career Development Award No. DA-00072. A preliminary version of this paper was presented at the 21st meeting of the NBER-NSF Seminar on Bayesian Inference in Econometrics, Chicago, October 31, 1980. We would like to thank H. Chernoff, R. Olshen, J. Koziol, D. Krantz, and R. Solow for reading and commenting on an earlier draft of this paper. The opinions and conclusions of this paper are the authors' sole responsibility.

REFERENCES

Casto, B.C., Hatch, G.C., Hsung, S.L., Huisingsh, J.L., Nesnow, S., and Waters, M.S. (1979), "Mutagenic and carcinogenic potency of extracts of diesel and related environmental emissions: in vitro mutagenesis and oncogenic transformation," Paper presented at the EPA International Symposium on the Health Effects of Diesel Engine Emissions. December 3-5, 1979. Cincinnati.

Cochran, W.G. (1980), "Summarizing the results of a series of experiments," Proceedings of the 25th Conference on Design of Experiments in Army Research, Development and Testing, ARD Report 80-2, Washington, D.C., U.S. Department of Defense, pp. 21-34.

Crouch, E. and Wilson, R. (1979), "Interspecies comparison of carcinogenic potency," Journal of Toxicology and Environmental Health, 5, 1095-1118.

——— and ——— (1980), "The regulation of carcinogens," Energy and Environmental Policy Center, Harvard University, Cambridge.

Dawid, A.P., Stone, M., and Zidek, J.V. (1973), "Marginalization paradoxes in Bayesian and structural inference (with Discussion)," Journal of the Royal Statistical Society, Ser. B, 35, 189-233.

De Finetti, B. (1964), "Foresight: its logical laws, its subjective sources," In Studies in Subjective Probability, eds. H.E. Kyburg, Jr and H.E. Smokler, New York: Wiley & Sons.

Dempster, A.P. (1980), "Comment on 'Using empirical Bayes techniques in the law school validity studies.' by D.B. Rubin," Journal of the American Statistical Association, 75, 817.

DuMouchel, W.H. (1981), "Documentation for CATDATA," Massachusetts Institute of Technology, Statistics Center, Technical Report No. , Cambridge.

Efron, B. and Morris, C. (1973), "Combining possibly related estimation problems (with Discussion)," *Journal of the Royal Statistical Society, Ser. B*, 35, 379-421.

Freireich, E.J., Gehan, E.A., Rall, D.P., Schmidt, L.H., and Skipper, H.E. (1966), "Quantitative comparison of toxicity of anti-cancer agents in mouse, rat, hamster, dog, monkey, and man," *Cancer Chemotherapy Reports*, 50, 219-245.

Giri, N. (1980), "Bayesian estimation of means of a two-way classification with random effect model," *Statistische Hefte*, 2, 254-260.

Hammond, E.C., Selikoff, I.J., Lawther, P.L., Seidman, H. (1976), "Inhalation of benzpyrene and cancer in man," *Annals of the New York Academy of Sciences*, 271, 116-124.

Harris, J.E. (1981), "Potential risk of lung cancer from diesel engine emissions," Washington: National Academy of Sciences.

Huisinigh, J.L., Bradow, R.L., Jungers, R.H., Harris, B.D., Zweidinger, R.B., Bushing, K.M., Gill, B.E., and Albert, R.E. (1979), "Mutagenic and carcinogenic potency of extracts of diesel and related environmental emissions: study design, sample generation, collection, and preparation," Paper presented at the EPA International Symposium on the Health Effects of Diesel Engine Emissions. December 3-5, 1979. Cincinnati.

Kadane, J.B., Dickey, J.M., Winkler, R.L. et al. (1980), "Interactive elicitation of opinion for a normal linear model," *Journal of the American Statistical Association*, 75, 845-854.

Kahn, H.A. (1966), "The Dorn study of smoking and mortality among U.S. veterans: report on eight and one-half years of observation," In *Epidemiological Approaches to the Study of Cancer and Other Chronic Disease*, ed. W. Haenszel, National Cancer Institute Monograph, 19, 1-125.

Lindley, D.V. and Smith, A.F.M. (1972), "Bayes estimates for the linear model (with Discussion)," Journal of the Royal Statistical Society, Ser. B, 34, 1-41.

Lloyd, W.J. (1971), "Long-term mortality study of steelworkers. V. Respiratory cancer in coke plant workers," Journal of Occupational Medicine, 13, 53-68.

Mazumdar, S., Redmond, C., Sollecito, W., Sussman, N. (1975), "An epidemiological study of exposure to coal tar pitch volatiles among coke oven workers," Journal of the Air Pollution Control Association, 25, 382-389.

Meselson, M. and Russell, K. (1977), "Comparisons of carcinogenic and mutagenic potency," In Origins of Human Cancer, eds. H.H. Hiatt, J.D. Watson, and J.A. Winsten, Cold Spring Harbor: Cold Spring Harbor Laboratory.

Mitchell, A.D., Evans, E.L., Jotz, M.M., Riccio, E.S., Mortelmans, K.E., and Simmon, V.F. (1979), "Mutagenic and carcinogenic potency of extracts of diesel and related environmental emissions: in vitro mutagenesis and DNA damage," Paper presented at the EPA International Symposium on the Health Effects of Diesel Engine Emissions. December 3-5, 1979. Cincinnati.

Nesnow, S., Triplett, L.L., Slaga, T.J. (1980), "Tumorigenesis of diesel exhaust, gasoline exhaust, and related emission extracts on Sencar mouse skin," Paper presented at the EPA Symposium on Application of Short-Term Bioassays in the Analysis of Complex Environmental Mixtures. March 4-7, 1980. Williamsburg.

Peto, R. (1977), "Epidemiology, multistage models, and short-term mutagenicity tests," In Origins of Human Cancer, eds. H.H. Hiatt, J. Watson, J. Winsten, Cold Spring Harbor: Cold Spring Harbor Laboratory.

Raiffa, H. and Schlaifer, R. (1968), Applied Statistical Decision Theory. Student Edition. Cambridge: The M.I.T. Press.

Schneiderman, M.A. (1967), "Mouse to man: statistical problems in bringing a drug to clinical trial," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, IV, 855-866.

Smith, A.F.M. (1973a), "Bayes estimates in one-way and two-way models," Biometrika, 60, 319-329.

————— (1973b), "A general Bayesian linear model," Journal of the Royal Statistical Society, Ser. B, 35, 67-75.

————— (1973c), "Discussion of 'Combining possibly related estimation problems.' by B. Efron and C. Morris," Journal of the Royal Statistical Society, Ser. B, 35, 410-412.

Tiao, G.C. and Zellner, A. (1964), "Bayes' theorem and the use of prior knowledge in regression analysis," Biometrika, 51, 119-130.

U.S. Environmental Protection Agency (1979), "The Carcinogen Assessment Group's final risk assessment on coke oven exposure," Washington. 38 pages.



Working Paper #278

never published

Date Due

renewed	FEB. 27 1997
MAR 6 '83	
NOV 14 '85	
FEB 16 '88	
APR 03 1992	
SEP. 5 1992	

Lib-26-67

ACME
BOOKBINDING CO., INC.

SEP 15 1983

100 CAMBRIDGE STREET
CHARLESTOWN, MASS.

MIT LIBRARIES



266

3 9080 004 415 912

MIT LIBRARIES

DUPL



267

3 9080 004 415 920

MIT LIBRARIES



268

3 9080 004 478 605

MIT LIBRARIES

DUPL



271

3 9080 004 446 859

MIT LIBRARIES

DUPL



275

3 9080 004 446 867

MIT LIBRARIES



276

3 9080 004 446 875

MIT LIBRARIES

DUPL



3 9080 004 446 883

MIT LIBRARIES



3 9080 004 478 571

